**Working paper**

**2019-05**

Statistics and Econometrics
ISSN 2387-0303

# Multivariate expectile trimming and the BExPlot

Ignacio Cascos, Maicol Ochoa

uc3m | Universidad Carlos III de Madrid

# Multivariate expectile trimming and the BExPlot

I. Cascos[1,2*] M. Ochoa[1]

[1]Deparment of Statistics, Universidad Carlos III de Madrid

[2]UC3M-BS Big Data Institute (IBiDat)

May 2019

**Abstract**

Expectiles are the solution to an asymmetric least squares minimization problem for univariate data. They resemble some similarities with the quantiles, and just like them, expectiles are indexed by a level $\alpha$. In the present paper, we introduce and discuss the main properties of the expectile multivariate trimmed regions, a nested family of sets, whose instance with trimming level $\alpha$ is built up by all points whose univariate projections lie between the expectiles of levels $\alpha$ and $1-\alpha$ of the projected dataset. Such trimming level is interpreted as the degree of centrality of a point with respect to a multivariate distribution and therefore serves as a depth function. We study here the convergence of the sample expectile trimmed regions to the population ones and the uniform consistency of the sample expectile depth. We also provide efficient algorithms for determining the extreme points of the expectile regions as well as for computing the depth of a point in $\mathbb{R}^2$. These routines are based on circular sequence constructions. Finally, we present some real data examples for which the Bivariate Expectile Plot (BExPlot) is introduced.

**Keywords** Algorithms, Bagplot, Data depth, Expectile, Trimmed regions

*Corresponding author: ignacio.cascos@uc3m.es

# 1 Introduction

*Expectiles* were first introduced by Newey and Powell (1987) in the context of linear regression as the solution to a minimization problem. They were so named because they resemble some similarities with the quan*tiles* of a random variable but unlike them, they are based on quadratic loss functions, as it is the case of the *expec*tation. They have received high attention in different areas such as risk measurement, due to the fact that some of them happen to be *elicitable* and *coherent risk measures*, as it was shown by Gneiting (2011), Bellini *et al.* (2014), and Ziegel (2016).

Inspired by Tukey (1975), who employed the quantiles of the univariate projections of a bivariate data cloud to produce a family of central (depth-trimmed) regions, Eilers (2010), and later on Giorgi and McNeil (2016), suggested to use the expectiles to build the *expectile regions* of a multivariate data set as the intersection of the halfspaces whose supporting hyperspace is determined by the expectile of a univariate projection of the data. In fact, the support functions of the expectile trimmed regions are given by the expectiles of the project dataset. Just like Tukey's central regions inspired the definition of the most popular depth function, the so-called halfspace depth, the expectile regions induce the *expectile depth*.

Based on the extensive studies of *depth functions*, their properties, and applications developed by Liu *et al.* (1999), Zuo and Serfling (2000), Dyckerhoff (2004), or Cascos (2010), we present here the expectile depth and discuss its main properties. Together with the expectile depth, we propose a novel exploratory data analytic tool, called the *Bivariate Expectile Plot* (BExPlot), which can be used for data visualization as an alternative to the *Bagplot*, see Rousseeuw *et al.* (1999).

Other multivariate generalizations of the expectiles have been considered by Breckling and Chambers (1988), who proposed the so-called multivariate $M$-quantiles as the solution to minimization problems similar to those giving rise to quantiles and expectiles. Since such $M$-quantiles may happen to lie out of the convex hull of the dataset they were built from, in

a posterior paper Breckling *et al.* (2001) presented an alternative definition of multivariate $M$-quantiles which still lack to be equivariant under arbitrary affine transformations. Later on, Maume-Deschamps *et al.* (2016) adapted the aforementioned notion of *elicitability* to the multivariate setting by using several vector and matrix norms. This way, they introduced the so-called *euclidean* (vector-valued) and *matrix* expectiles. More recently, Herrmann *et al.* (2018) introduced another multivariate version of expectiles, the *geometric* expectiles, as the unique solution of a convex risk minimization problem. The homogeneity and equivariance under orthogonal transformations properties of geometric expectiles makes them appealing multivariate risk measures.

The paper is organized as follows: in Section 2 we review the concept of (univariate) expectile of a random variable, while Section 3 is devoted to multivariate expectile regions. Specifically, we introduce the expectile regions together with their meain properties and characterise the extreme points of the sample expectile regions. In Section 4 we present the expectile depth and discuss how to compute its 2-dimensional empirical version. In Section 5 we study the consistency of the sample expectile regions with respect to the Hausdorff metric and the uniform consistency of the sample expectile depth function. Finally, in Section 6 we introduce the Bivariate Expectile Plot as an EDA tool, while the main contributions of the paper are highlighted in Section 7. Two appendices are placed at the end of the manuscript, the first of them describes two algorithms to compute the bivariate expectile regions and depth, while the second contains the proofs of some mathematical results.

# 2 Univariate expectiles

Given a random variable $X$ defined on a general probability space and with finite first moment, $\mathbb{E}|X| < \infty$, and given $\alpha \in (0,1)$, the $\alpha$-*expectile* of $X$ was defined in Newey and Powell (1987) as the minimizer of the quadratic expression

$$e_\alpha(X) := \arg\min_x \left\{(1-\alpha)\mathbb{E}(X-x)_-^2 + \alpha\mathbb{E}(X-x)_+^2\right\}, \tag{1}$$

where $a_+ = \max\{a, 0\}$ and $a_- = -\min\{a, 0\}$ for any $a \in \mathbb{R}$. Observe that the $\alpha$-quantile of $X$ is the minimizer of an expression that can be obtained from (1) deleting the squares.

Considering the first order condition obtained from (1), it is not hard to see that $e_\alpha(X)$ is the unique solution to the equation

$$-(1-\alpha)\mathbb{E}(X - e_\alpha(X))_- + \alpha\mathbb{E}(X - e_\alpha(X))_+ = 0\,. \tag{2}$$

The expressions below, which will turn out to be useful when interpreting and computing expectiles, can be obtained from (2) after some elementary algebra

$$e_\alpha(X) = \mathbb{E}[X] + \frac{2\alpha - 1}{1-\alpha}\mathbb{E}(X - e_\alpha(X))_+ \tag{3}$$

$$= \frac{(1-\alpha)\int_0^{F_X(e_\alpha(X))} F_X^{-1}(t)\,\mathrm{d}t + \alpha\int_{F_X(e_\alpha(X))}^1 F_X^{-1}(t)\,\mathrm{d}t}{\alpha + (1-2\alpha)F_X(e_\alpha(X))}\,, \tag{4}$$

where $F_X$ stands for the cdf of $X$, while $F_X^{-1}$ is its quantile function.

**Fundamental geometric interpretation of the expectiles** At the light of the Choquet integral of random variables $(X - e_\alpha(X))_-$ and $(X - e_\alpha(X))_+$, it follows from (2) that the $\alpha$-expectile of $X$ is the point such that the area between the graph of cdf of $X$ and the horizontal axis and to left of $e_\alpha(X)$ is exactly $\alpha/(1-\alpha)$ times the area between the cdf of $X$ and horizontal line $y = 1$ and to the right of $e_\alpha(X)$, that is,

$$(1-\alpha)\int_{-\infty}^{e_\alpha(X)} F_X(x)\,\mathrm{d}x = \alpha\int_{e_\alpha(X)}^{+\infty}(1 - F_X(x))\,\mathrm{d}x\,.$$

In an alternative but equivalent way, and as can be better seen in (4), the $\alpha$-expectile of a random variable is a convex combination of the gravity centres of its lower and upper tails with regard to itself, such that the weight of the lower tail gravity centre corresponds to the weight of the upper one in a proportion of $(1-\alpha)$ to $\alpha$.

See Figure 1 for the 0.75-expectile of a standard normal random variable $X$. The vertical line is $x = e_{0.75}(X) = 0.4363$ and the area of the shaded region to its left is $3 = 0.75/(1-0.75)$ times the area of the shaded region to its right.

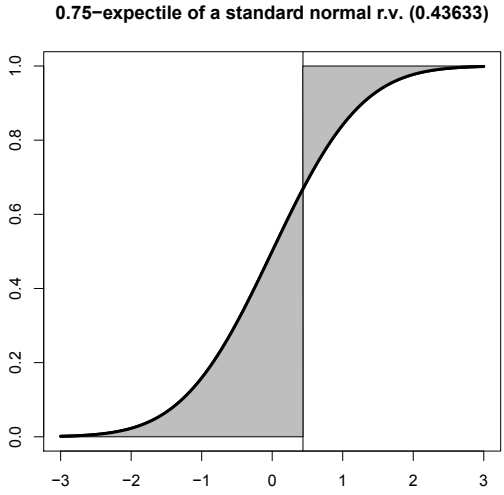



Figure 1: Geometric interpretation of the $\alpha$-expectile

We take advantage of the fundamental geometric interpretation of the expectiles to provide two further explanations.

1. *Gambler's interpretation.* If the monetary return of a gamble is modeled as $X$ and a gambler has to pay an amount $x$ in order to play, the gambler's gain will be given by $(X-x)_+$ while his loss is $(X-x)_-$. The $\alpha$-expectile of $X$ is the price to play when the expected gain is $(1-\alpha)/\alpha$ times the expected loss. In a fair gamble, the expected gain matches the expected loss, so the price to play is exactly $e_{1/2}(X) = \mathbb{E}[X]$.

   If a risk averse gambler refuses to play unless her expected gain is at least $a > 1$ times the expected loss, the price she would be willing to pay is the $(a+1)^{-1}$-expectile of $X$. On the other hand, a desperate gambler might agree to pay a price that makes the expected losses not more than $a > 1$ times the expected gains, thus she would agree to pay up to the $a/(a+1)$-expectile of $X$.

2. *Actuarial interpretation.* If $X$ models the final payment of an insurance company for one of its products at the moment it expires, the $\alpha$-expectile of $X$ is the premium

(excluding administrative expenses) so that the policyholder's expected saving (due to the acquisition of the insurance product) is $(1-\alpha)/\alpha$ times her expected expenditure, once the premium is deduced from the final payment.

## 2.1 Properties of expectiles

We list below the properties of univariate expectiles that will be needed to introduce the expectile regions and depth. Their proofs can be found in classical references, specifically Bellini *et al.* (2014).

1. *Most central expectile,* $e_{1/2}(X) = \mathbb{E}X$.
2. *Upper and lower expectiles,* $e_\alpha(X) = -e_{1-\alpha}(-X)$.
3. *Translation equivariance,* $e_\alpha(X+a) = a + e_\alpha(X)$ for $a \in \mathbb{R}$.
4. *Homogeneity,* $e_\alpha(\lambda X) = \lambda e_\alpha(X)$ for $\lambda \geq 0$.
5. *Monotonicity,* if $X \leq Y$ a.s., then $e_\alpha(X) \leq e_\alpha(Y)$.
6. *Subadditivity,* for $1/2 \leq \alpha < 1$, $e_\alpha(X+Y) \leq e_\alpha(X) + e_\alpha(Y)$.
7. *Superadditivity,* for $0 \leq \alpha < 1/2$, $e_\alpha(X+Y) \geq e_\alpha(X) + e_\alpha(Y)$.
8. *Parameter continuity,* $e_\alpha$ is continuous in $\alpha$.
9. *Strict parameter monotonicity,* $e_\alpha$ is strictly increasing in parameter $\alpha$.

## 2.2 Some specific expectiles

In order to obtain some insight about expectiles, we present next the expectiles of some standard distribution models.

**Bernoulli random variable** If $X$ is a random variable such that $P(X = 1) = p$, and $P(X = 0) = 1 - p$, then,

$$e_\alpha(X) = \frac{\alpha p}{1 - \alpha - p + 2\alpha p}.$$

**Uniform random variable** If $X$ is a uniform random variable on $(0, 1)$, then

$$e_\alpha(X) = \frac{\alpha - \sqrt{\alpha(1-\alpha)}}{2\alpha - 1},$$

if $\alpha \neq 1/2$, while $e_{0.5}(X) = 1/2$.

**Gaussian random variable** The $\alpha$-expectile of a univariate standard normal random variable $X$ is the unique solution to the equation

$$\frac{\phi(e_\alpha(X))}{e_\alpha(X)} + \Phi(e_\alpha(X)) = \frac{\alpha}{2\alpha - 1}, \tag{5}$$

where $\phi$ is the standard normal density and $\Phi$ the standard normal cdf, if $\alpha \neq 1/2$, while $e_{0.5}(X) = 0$. See Figure 2 for the graphs of the expectile and quantile functions of a standard normal random variable.

**Inverse expectile function** The same way the expectile function plays a similar role to the quantile function, the inverse expectile function is somehow analogous to the cumulative distribution function (inverse quantile). This inverse expectile function will be a helpful tool when we study the *expectile depth* in Section 4.

Any real value inside the interior of the convex hull of the support of $X$ corresponds to one of its expectiles. After some elementary algebraic transformations in (2), it is straightfoward

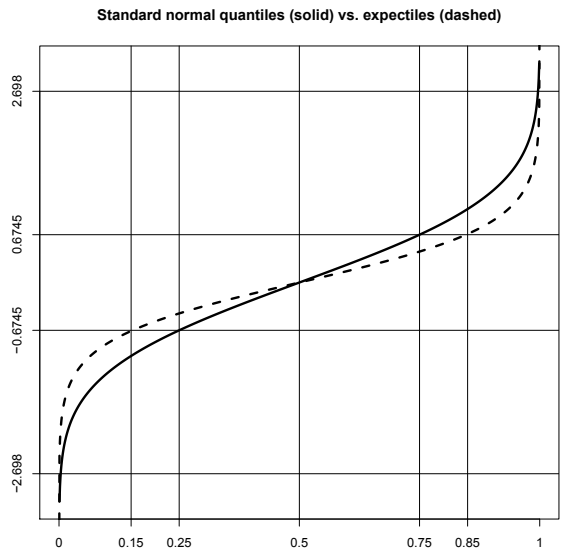



Figure 2: The 0.15-expectile (0.85-expectile) of a normal random variable approximately coincides with its 0.25-quantile (0.75-quantile).

to see that any $\operatorname{ess\,inf}(X) < x < \operatorname{ess\,sup}(X)$ is the $e_X^{-1}(x)$-expectile of $X$ with

$$\begin{aligned} e_X^{-1}(x) &= \left(1 + \frac{\mathbb{E}(X-x)_+}{\mathbb{E}(X-x)_-}\right)^{-1} \\ &= \frac{\mathbb{E}(X-x)_-}{\mathbb{E}|X-x|} \\ &= \frac{x - \mathbb{E}X + \mathbb{E}(X-x)_+}{x - \mathbb{E}X + 2\mathbb{E}(X-x)_+}. \end{aligned} \tag{6}$$

Due to the relation between upper and lower expectiles (property 2. in Section 2.1), the inverse expectile satisfies $e_{-X}^{-1}(-x) = 1 - e_X^{-1}(x)$. Moreover by the strict parameter monotonicity (property 9. in Section 2.1), the inverse expectile function is strictly increasing.

**Characterization property** The expectiles of a random variable (with finite first moment) characterize its distribution. In fact, after the relation between upper and lower expectiles (property 2. in Section 2.1), only either the upper or lower expectiles are needed to characterize the distribution. This is easily verified from the inverse expectile function

$e_X^{-1}$ which, as seen in (6), determines $\mathbb{E}(X-x)_+$, the stop-loss function associated with $X$. After the relation

$$\mathbb{E}(X-x)_+ = \int_x^\infty (1-F_X(t))\,\mathrm{d}t$$

the stop-loss function is sometimes called integrated survival function of a random variable, see Müller and Stoyan (2002), and immediately characterizes its distribution.

### 2.3 Sample expectiles

Consider $x_1, x_2, \ldots, x_n \in \mathbb{R}$ a sample of univariate observations and $0 < \alpha < 1$. The empirical $\alpha$-expectile, denoted by $e_\alpha(x_1,\ldots,x_n)$ or shortly $e_{n,\alpha}$, is the solution to equation (3), which when written in the form of weighted sums adopts the expression

$$\begin{aligned} e_{n,\alpha} &= \overline{x} + \frac{2\alpha-1}{n(1-\alpha)} \sum_{x_i > e_{n,\alpha}} (x_i - e_{n,\alpha}) \\ &= \frac{1-\alpha}{\alpha n + nF_n(e_{n,\alpha})(1-2\alpha)} \left( n\overline{x} + \frac{2\alpha-1}{1-\alpha} \sum_{x_i > e_{n,\alpha}} x_i \right), \end{aligned} \tag{7}$$

where, as usual, $\overline{x}$ stands for the sample mean, and $F_n(\cdot)$ denotes the empirical cdf.

Fortunately enough, $e_{n,\alpha}$ can be computed fast using a Newton-gradient algorithm implemented as a built-in function in the R package `expectreg`, see Sobotka *et al.* (2014).

Let $\min x_i \le x \le \max x_i$, according to (6), in order to find out to which sample expectile does $x$ correspond to, we only need to compute

$$e_n^{-1}(x) = \frac{nx - n\overline{x} + \sum_{x_i > x}(x_i - x)}{nx - n\overline{x} + 2\sum_{x_i > x}(x_i - x)}\,. \tag{8}$$

## 3 Expectile trimmend regions

Following Eilers (2010) and the scenario set construction of Giorgi and McNeil (2016), for any random vector $\boldsymbol{X}$ with finite first moment $\mathbb{E}\|\boldsymbol{X}\| < \infty$, we define its *expectile trimmed*

*region* of level $0 < \alpha \leq 1/2$, denoted by $\mathrm{E}^{\alpha}(\boldsymbol{X})$, as an intersection of closed halfspaces supported by hyperplanes whose constant term is determined by a univariate expectile

$$\mathrm{E}^{\alpha}(\boldsymbol{X}) := \bigcap_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{u}\rangle \leq e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)\}\,, \tag{9}$$

where $\mathbb{S}^{d-1}$ stands for the unit sphere in $\mathbb{R}^d$ and $\langle \cdot, \cdot \rangle$ represents the standard inner product in $\mathbb{R}^d$.

Notice that for $d = 1$, the expectile region is the closed interval of real numbers

$$\mathrm{E}^{\alpha}(X) = [e_{\alpha}(X), e_{1-\alpha}(X)]\,.$$

As an intersection of closed halfspaces, the expectile regions are closed convex sets, and thus characterized by means of their support functions, see e.g. Schneider (1993). The *support function* of a closed convex non-empty set $K \subset \mathbb{R}^d$ evaluated at $\boldsymbol{u} \in \mathbb{R}^d$ is defined as

$$h(K, \boldsymbol{u}) := \sup\{\langle \boldsymbol{x}, \boldsymbol{u}\rangle : \boldsymbol{x} \in K\}\,.$$

If $h(K, \cdot)$ is a support function, then it is clearly positively homogeneous and subadditive, while the reverse does also hold as can be found in Schneider (1993, Th.1.7.1): given any positively homogeneous and subadditive function $h : \mathbb{R}^d \to \mathbb{R}$ there exists a unique compact and convex set $K \subset \mathbb{R}^d$ such that $h$ is its support function.

From the homogeneity of subadditivity of univariate expectiles (properties 4. and 6. in Section 2.1), we know that for any $d$-dimensional random vector with finite first moment $\boldsymbol{X}$ and $0 < \alpha \leq 1/2$ the map $\boldsymbol{u} \mapsto e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is positively homogeneous and subadditive, so it constitutes the support of a compact and convex set, which is in fact $\mathrm{E}^{\alpha}(\boldsymbol{X})$. Hence, alternatively to (9), the expectile trimmed regions can be characterized in terms of its support function as

$$h(\mathrm{E}^{\alpha}(\boldsymbol{X}), \boldsymbol{u}) = \sup\{\langle \boldsymbol{x}, \boldsymbol{u}\rangle : \boldsymbol{x} \in \mathrm{E}^{\alpha}(\boldsymbol{X})\} = e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)\,. \tag{10}$$

An important consequence of (10) is that, since $e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is the upper end-point of the expectile region of $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ of level $\alpha$, the expectile depth satisfies the strong (also weak) projection property, see Dyckerhoff (2004, Th.3).

## 3.1 Properties of the expectile trimmed regions

The expectile regions fulfill the usual properties for depth-trimmed regions considered in Dyckerhoff (2004) and Cascos (2010), that is, if $0 < \alpha \leq 1/2$ then it is not hard to see that

1. *Most central point,* $\mathrm{E}^{1/2}(\boldsymbol{X}) = \{\mathbb{E}\boldsymbol{X}\}$.
2. *Nesting,* $\mathrm{E}^{\alpha}(\boldsymbol{X}) \subset \mathrm{E}^{\beta}(\boldsymbol{X})$ if $0 < \beta \leq \alpha \leq 1/2$.
3. *Convexity,* $\mathrm{E}^{\alpha}(\boldsymbol{X})$ is convex.
4. *Compactness,* $\mathrm{E}^{\alpha}(\boldsymbol{X})$ is compact.
5. *Affine equivariance,* $\mathrm{E}^{\alpha}(A\boldsymbol{X}+\boldsymbol{b}) = A\mathrm{E}^{\alpha}(\boldsymbol{X})+\boldsymbol{b}$ for any matrix $A \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$.
6. *Monotonicity,* if $\boldsymbol{X} \leq \boldsymbol{Y}$ (componentwisely) a.s., then $\mathrm{E}^{\alpha}(\boldsymbol{Y}) \oplus \mathbb{R}^d_+ \subseteq \mathrm{E}^{\alpha}(\boldsymbol{X}) \oplus \mathbb{R}^d_+$.
7. *Minkowski subadditivity,* $\mathrm{E}^{\alpha}(\boldsymbol{X}+\boldsymbol{Y}) \subseteq \mathrm{E}^{\alpha}(\boldsymbol{X}) \oplus \mathrm{E}^{\alpha}(\boldsymbol{Y})$,

where the symbol $\oplus$ stands for the Minkowski or elementwise set addition, i.e., given $K_1, K_2$ two subsets of $\mathbb{R}^d$, we have $K_1 \oplus K_2 = \{\boldsymbol{x}+\boldsymbol{y} : \boldsymbol{x} \in K_1, \boldsymbol{y} \in K_2\}$.

**Characterization property** For any random vector with finite first moment $\boldsymbol{X}$, $\boldsymbol{u} \in \mathbb{R}^d$, and $0 < \alpha \leq 1/2$, the support function of $\mathrm{E}^{\alpha}(\boldsymbol{X})$ evaluated on $\boldsymbol{u}$ is the $(1-\alpha)$-expectile of the linear combination of its components $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$, see (10). After the characterization property of the upper expectiles, the distribution of every $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ is thus determined by $\{\mathrm{E}^{\alpha}(\boldsymbol{X})\}_{0<\alpha\leq 1/2}$. Finally, and according to Cramér and Wold (1936), the family of all expectile regions of $\boldsymbol{X}$ characterizes its distribution.

**Continuity properties of the expectile regions** In order to discuss some continuity property for the expectile regions with respect to the trimming level $\alpha$, we need to introduce a notion of distance for compact and convex sets. If $K_1, K_2 \subset \mathbb{R}^d$ are compact and convex

with support functions $h(K_1, \cdot)$ and $h(K_2, \cdot)$, the *Hausdorff distance* between $K_1, K_2$, see Molchanov (2017, H.5), is

$$d_H(K_1, K_2) := \sup_{\boldsymbol{u} \in \mathbb{S}^{d-1}} |h(K_1, \boldsymbol{u}) - h(K_2, \boldsymbol{u})| .$$

**Proposition 3.1.** *If* $\lim_n \alpha_n = \alpha$ *in* $(0, 1/2]$, *then*

$$\lim_n d_H(\mathrm{E}^{\alpha_n}(\boldsymbol{X}), \mathrm{E}^{\alpha}(\boldsymbol{X})) = 0 .$$

See Appendix B.1 for the proof of Proposition 3.1.

In Section 4 we study the notion of depth induced by the expectile regions. Proposition 3.2 below about expectile regions together with Dyckerhoff (2017, Th.3.1) guarantee that such depth is continuous.

**Proposition 3.2.** *If* $\alpha > \beta$ *in* $(0, 1/2]$ *then*

$$\mathrm{E}^{\alpha}(\boldsymbol{X}) \subset \operatorname{int} \mathrm{E}^{\beta}(\boldsymbol{X}) .$$

See Appendix B.2 for the proof of Proposition 3.2.

## 3.2 Sample expectile trimmed regions

For a sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $0 \leq \alpha < 1/2$, the sample $\alpha$-expectile region is the set $\mathrm{E}_n^{\alpha}$ whose support function, see (10), is

$$h(\mathrm{E}_n^{\alpha}, \boldsymbol{u}) = e_{n,1-\alpha}(\langle \boldsymbol{x}_1, \boldsymbol{u} \rangle, \ldots, \langle \boldsymbol{x}_n, \boldsymbol{u} \rangle) ,$$

that is, it corresponds to the $(1-\alpha)$-expectile of the (univariate) sample $\langle \boldsymbol{x}_1, \boldsymbol{u} \rangle, \ldots, \langle \boldsymbol{x}_n, \boldsymbol{u} \rangle$.

Finally, according to (7), the sample $\alpha$-expectile trimmed region can be written as the convex hull of all linear combinations of the points from the data sample with some prescribed weights

$$\mathrm{E}_n^{\alpha} = \mathrm{co} \left\{ \boldsymbol{x} = \Lambda_{n,s}^{\alpha} \left( n\overline{\boldsymbol{x}} + \frac{1-2\alpha}{\alpha} \sum_{\pi_{\boldsymbol{u}}(i) > s} \boldsymbol{x}_i \right) : \boldsymbol{u} \in \mathbb{S}^{d-1} \text{ and } G_{\boldsymbol{u}}(\boldsymbol{x}) = s \right\} , \qquad (11)$$

where co stands for the convex hull, $\Lambda_{n,s}^{\alpha} = \alpha[(1-\alpha)n + s(2\alpha-1)]^{-1}$, $\pi_{\boldsymbol{u}}$ is the permutation from $\{1, 2, \ldots, n\}$ such that $\langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(1)}, \boldsymbol{u}\rangle \leq \ldots \leq \langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(n)}, \boldsymbol{u}\rangle$, and $G_{\boldsymbol{u}}(\boldsymbol{x}) = s$ if and only if $\langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(s)}, \boldsymbol{u}\rangle \leq \langle \boldsymbol{x}, \boldsymbol{u}\rangle \leq \langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(s+1)}, \boldsymbol{u}\rangle$.

Notice that for any $\boldsymbol{u} \in \mathbb{S}^{d-1}$ there is, at least, one extreme point of $\mathrm{E}_n^{\alpha}$ in the supporting hyperplane $\{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{u}\rangle = e_{n,1-\alpha}(\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle, \ldots, \langle \boldsymbol{x}_n, \boldsymbol{u}\rangle)\}$. In Appendix A.1 we present an algorithm to compute all the extreme points of the expectile region of level $\alpha$ with respect to a bivariate dataset and discuss its complexity.

# 4 Expectile depth function

The *expectile depth* associates to each point $\boldsymbol{y} \in \mathbb{R}^d$ its degree of centrality with respect to the distribution of the random vector $\boldsymbol{X}$ in terms of the expectile regions. Specifically, the expectile depth of a point is the supremum all levels $\alpha$ such that the given point belongs to the expectile region of level $\alpha$,

$$\mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) := \sup\{0 < \alpha \leq 1/2 : \boldsymbol{y} \in \mathrm{E}^{\alpha}(\boldsymbol{X})\}.$$

Conversely, the expectile region of level $\alpha$ consists of all the points whose depth is at least $\alpha$, $\mathrm{E}^{\alpha}(\boldsymbol{X}) = \{\boldsymbol{y} : \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) \geq \alpha\}$, and after the weak projection property, the expectile depth of a point with respect to a random vector can be computed as the infimum of the depths of the univariate projections of the point w.r.t. the projected random vector, that is,

$$\mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) = \inf_{\boldsymbol{u} \in \mathbb{S}^{d-1}} \mathrm{ED}(\langle \boldsymbol{y}, \boldsymbol{u}\rangle; \langle \boldsymbol{X}, \boldsymbol{u}\rangle).$$

An immediate consequence is that the expectile depth with respect to a random variable $X$ (for $d = 1$) is written in terms of the inverse expectile as $\mathrm{ED}(y; X) = \min\{e_X^{-1}(y), e_{-X}^{-1}(-y)\}$.

The expression for the expectile depth derived in Proposition 4.1 below will turn out to be crucial when explicitly computing empirical expectile depths.

**Proposition 4.1.** *For any $\boldsymbol{y} \in \mathbb{R}^d$, the expectile depth function satisfies*

$$\mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) = \left(2 - \inf_{\boldsymbol{u} \in \mathbb{S}^{d-1}} \frac{\langle \mathbb{E}\boldsymbol{X} - \boldsymbol{y}, \boldsymbol{u} \rangle}{\mathbb{E}\langle \boldsymbol{X} - \boldsymbol{y}, \boldsymbol{u} \rangle_+}\right)^{-1} .$$

See Appendix B.3 for the proof of Proposition 4.1.

## 4.1 Properties of the expectile depth

The expectile depth function satisfies the following properties that can be immediately derived from those of the expectile regions presented in Section 3.1.

1. *Affine invariance,* the depth of a point $\boldsymbol{y} \in \mathbb{R}^d$ is independent of the coordinate system. For any matrix $A \in \mathbb{R}^{d \times d}$ and $\boldsymbol{b} \in \mathbb{R}^d$

$$\mathrm{ED}(A\boldsymbol{y} + \boldsymbol{b}; A\boldsymbol{X} + \boldsymbol{b}) = \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) .$$

2. *Continuity,* the mapping $\boldsymbol{y} \mapsto \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X})$ is continuous by Proposition 3.2 and Dyckerhoff (2017, Th.3.1).

3. *Maximality at centre,* the expectile depth attains its unique maximum at the mean, $\mathbb{E}\boldsymbol{X}$, in fact $\mathrm{ED}(\mathbb{E}\boldsymbol{X}; \boldsymbol{X}) = 1/2$.

4. *Quasiconcavity,* as a consequence of the convexity of the expectile regions,

$$\mathrm{ED}(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}; \boldsymbol{X}) \geq \min\{\mathrm{ED}(\boldsymbol{x}; \boldsymbol{X}), \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X})\} \quad \text{for } 0 \leq \lambda \leq 1 .$$

5. *Vanishing at infinity.* The expectile depth of a point $\boldsymbol{y}$ goes to zero as $\|\boldsymbol{y}\| \to \infty$.

Furthermore, the expectile depth is strictly monotone in the sense of Dyckerhoff (2017).

**Proposition 4.2.** *Strict monotonicity of the expectile depth.*

*a) The expectile depth is strictly monotone in rays from the center of the distribution.*

*b) For $0 < \alpha < 1/2$ it holds $\mathrm{E}^\alpha(\boldsymbol{X}) = \mathrm{cl}\{\boldsymbol{y} : \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) > \alpha\}$.*

See Appendix B.4 for the proof of Proposition 4.2.

## 4.2 Sample expectile depth function

Our goal is now to define the empirical expectile depth function w.r.t. to the sample $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$. To simplify the computations, we assume that the point $\boldsymbol{y}$ whose depth is to be assessed is the origin of coordinates, in any other case, the depth is computed after cetering the sample at the given point. The infimum in Proposition 4.1 is now

$$\inf_u \frac{\sum \langle \boldsymbol{x}_i, \boldsymbol{u} \rangle}{\sum \langle \boldsymbol{x}_i, \boldsymbol{u} \rangle_+},$$

and the *sample expectile depth* of the origin can thus be written as

$$\mathrm{ED}_n(\mathbf{0}) = \left(2 - \inf_u \frac{\sum \langle \boldsymbol{x}_i, \boldsymbol{u} \rangle}{\sum \langle \boldsymbol{x}_i, \boldsymbol{u} \rangle_+}\right)^{-1}. \tag{12}$$

In $\mathbb{R}^2$ we have $\boldsymbol{u} \in \mathbb{S}^1$, and thus we can write $\boldsymbol{u} = (\cos\gamma, \sin\gamma)$ for some $\gamma \in [0, \pi)$. We can further assume that the sample average, $\overline{\boldsymbol{x}}$, lies on the negative part of the vertical axis (rotate the sample if needed). Therefore, the infimum in $\mathrm{ED}_n(\mathbf{0})$ turns into

$$\begin{aligned}
\inf_{\boldsymbol{u}} \frac{\sum \langle \boldsymbol{x}_i, \boldsymbol{u} \rangle}{\sum \langle \boldsymbol{x}_i, \boldsymbol{u} \rangle_+} &= \inf_{\gamma} \frac{\sum \langle \boldsymbol{x}_i, (\cos\gamma, \sin\gamma) \rangle}{\sum \langle \boldsymbol{x}_i, (\cos\gamma, \sin\gamma) \rangle_+} \\
&= \min_i \frac{\|\overline{\boldsymbol{x}}\|}{\|\overline{\boldsymbol{x}}(\gamma)\|} \inf \left\{ \frac{\langle (0,-1), (\cos\gamma, \sin\gamma) \rangle}{\langle (\cos\beta, \sin\beta), (\cos\gamma, \sin\gamma) \rangle_+} : \gamma_i + \pi/2 < \gamma < \gamma_{i+1} + \pi/2 \right\} \\
&= \min_i \frac{\|\overline{\boldsymbol{x}}\|}{\|\overline{\boldsymbol{x}}(\gamma)\|} \inf \left\{ \frac{-\sin\gamma}{\cos(\gamma - \beta)} : \gamma_i + \pi/2 < \gamma < \gamma_{i+1} + \pi/2 \right\},
\end{aligned}$$

where $\gamma_i$ is the angle between the line through the origin and point $\boldsymbol{x}_i$ with respect to the positive horizontal axis and $\overline{\boldsymbol{x}}(\gamma) = \|\overline{\boldsymbol{x}}(\gamma)\|(\cos\beta, \sin\beta)$ is the average of the sample points in the halfspace with inner normal $(\cos(\gamma + \pi/2), \sin(\gamma + \pi/2))$.

As a standard constrained minimization procedure, we deduce the optimal solution at each circular segment from the KKT conditions:

$$\begin{aligned}
&\text{if } \frac{\pi}{2} < \beta < \frac{3\pi}{2} \text{ then the solution is } \gamma^* = \gamma_i + \frac{\pi}{2}, \\
&\text{if } -\frac{\pi}{2} < \beta < \frac{\pi}{2} \text{ then the solution is } \gamma^* = \gamma_{i+1} + \frac{\pi}{2}.
\end{aligned} \tag{13}$$

In Appendix A.2 we present an efficient algorithm for computing the depth of a point w.r.t. a bivariate sample, $d = 2$.

# 5 Consistency of the expectile regions and depth

We establish here the convergence of the sample expectile trimmed regions $\{\mathrm{E}_n^\alpha : n \geq 1\}$ to the population one $\mathrm{E}^\alpha(\boldsymbol{X})$ with respect to the Hausdorff metric. Furthermore, we show that the sample expectile depth function is a uniformly consistent estimator of the population expectile depth function.

**Theorem 5.1.** *Given any d-dimensional random vector $\boldsymbol{X}$ with finite first moment, the family of sample expectile regions $\{\mathrm{E}_n^\alpha\}_\alpha$ and the sample expectile depth function $\mathrm{ED}_n$ built from a random sample of $\boldsymbol{X}$ satisfy:*

*a) For any compact set $I \subset (0, 1/2)$ we have*

$$\sup_{\alpha \in I} d_H(\mathrm{E}_n^\alpha, \mathrm{E}^\alpha(\boldsymbol{X})) \xrightarrow{a.s.} 0\,.$$

*b) It holds*

$$\sup_{\boldsymbol{y} \in \mathbb{R}^d} |\mathrm{ED}_n(\boldsymbol{y}) - \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X})| \xrightarrow{a.s.} 0\,.$$

See Appendix B.5 for the proof of Theorem 5.1.

# 6 The Bivariate Expectile Plot (BExPlot)

The *BExPlot* is a graphical representation for bivariate data inspired by the *bagplot*, see Rousseeuw *et al.* (1999). It consists of a bullet located at the sample mean (singleton representing the 1/2-expectile region), a shaded bag representing the 0.15-expectile region, and a convex fence enclosing all data points that are contained inside the bag after expanding it from the sample mean by a factor of 4. The points outside the fence (if any) are marked as

possible outliers. As we present in Figure 3 left, the BExPlot can be plotted with *expectile-plots* for each of the variables at margins.

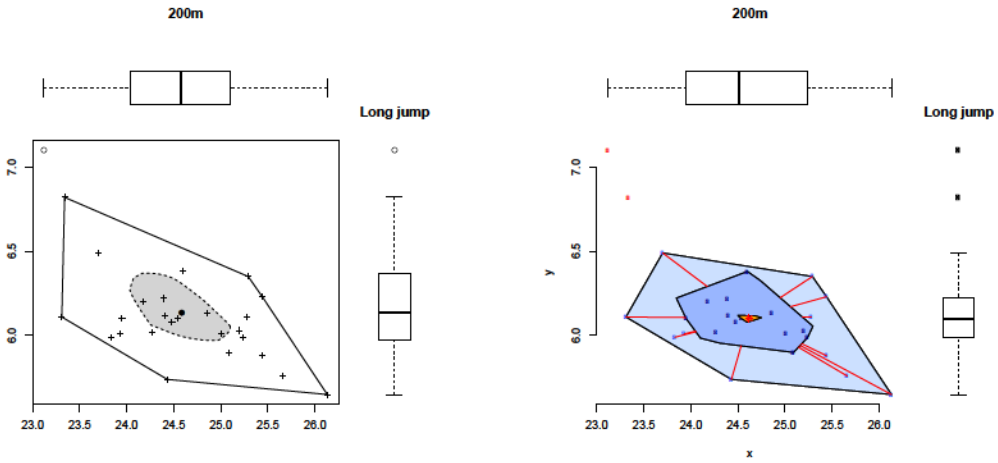
Given a univariate dataset, the expectile-plot is similar to a boxplot. It represents the minimum and maximum at the whiskers, contains a box raging from the 0.15-expectile to the 0.85-expectile, and a mark on the average value (0.5-expectile). Observe that, as seen in Figure 2, the 0.15- and 0.85-expectiles of a standard normal distribution approximately coincide with its first and third quartiles, while the mean and the median also coincide, resulting in an expectile-plot very similar to a boxplot when the underlying distribution is normal. As in a boxplot, possible outliers are those points located farther than 1.5 times the width of the box (distance between the 0.15- and the 0.85-expectiles) from the end-points of the box.

All points marked as possible outliers in the expectile-plot of one of the marginals, will be marked as possible outliers in the BExPlot. Any other point marked as possible outlier in the BExPlot will correspond to an observation that would have been marked as possible outlier in the expectile-plot obtained for some univariate projection of the dataset.

## 6.1 Single BExPlot

In Figure 3 left we have represented a BExPlot of time in the 200m race (horizontal axis) and the distance of the long jump (vertical axis) for the 30 athletes that took part in the Decathlon of the 1992 Barcelona Olympic Games. At margins we represent the expectile-plots of the times at the 200m race and the long jump. The shaded region in the BExPlot is the 0.15-expectile region. Only one point is out of the fence and appears marked as possible outlier. It corresponds to an athlete marked as possible outlier in the long jump expectile-plot.

The bagplot of the Barcelona92 dataset plotted with Wolf (2018), is represented in Figure 3 right. In this chart, two points are marked as possible outliers and the plots at margins




(a) BExPlot (b) Bagplot

Figure 3: BExPlot and Bagplot for Barcelona1992

are the boxplots of the 200m race and the long jump.

## 6.2 Multiple BExPlots

In order to represent the BExPlots of each pair of more than two variables we use a matrix-type chart where each cell contains the BExPlot of the variables that cross at it, while expectile-plots are represented at margins. Figure 4 is built out of Fisher's classical Iris dataset, specifically variables sepal length, sepal width, and petal length of species virginica are considered. It presents a matrix of BExPlots with expectile-plots at margins.

Notice that some of the possible outliers detected in some of the BExPlots in Figure 4 are not detected in any of the two expectile-plots at its margins. This is because they correspond to possible outliers that would have been detected on univariate projections different from the marginals.

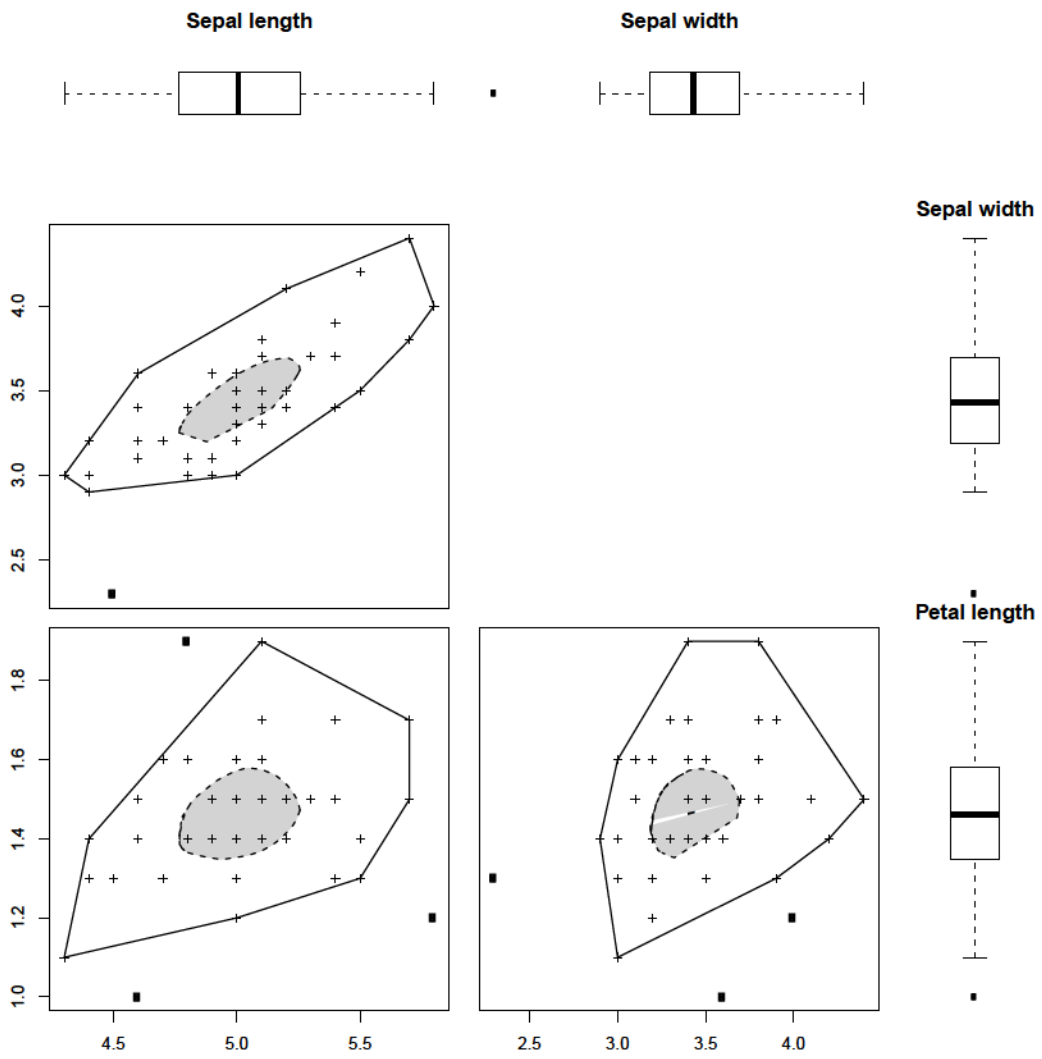



Figure 4: Multiple BExPlots

## 6.3 BExPlots with confidence regions

Finally we show an approximation of the confidence region on the mean of a bivariate dataset based on expectile regions. At margins an approximation of the 95% confidence interval on the mean of each variable is represented in a notched expectile-plot.

For any dimension $d$, the level $\alpha$ such that the $\alpha$-expectile region of a $d$-dimensional standard normal distribution matches the region centred at the origin and containing the

sample mean with probability $\beta$ is computed. The empirical expectile region of such level $\alpha$ will serve as an approximation of a confidence region on the mean with confidence level $\beta$.

After some elementary algebra, the trimming level is the solution, on $\alpha$, to an equation built from (5) substituting $e_\alpha(X)$ by $-\sqrt{F^{-1}_{\chi^2_d}(\beta)/n}$, where $F^{-1}_{\chi^2_d}$ is the quantile function of a chi-squared random variable with $d$ degrees of freedom. For $d = 1$ this expression is used to compute the $\alpha$ that determines the notches, while for $d = 2$ the expression is used to compute the $\alpha$ that determines the approximation to the confidence region presented together with the BExPlot.

In Figure 5 we present a BExPlot of the *Hemophilia* dataset from Pokotylo *et al.* (2019). The dataset contains data of AHF activity (variable $x$) and AHF antigen (variable $y$) on the blood of two groups of women, 45 of them being Hemophilia A carriers (marked as + in the chart) and 30 being non-carriers of Hemophilia A (marked as $\times$ in the chart).

The notches in the expectile-plots at margins represent an approximation of the 95% confidence interval on each mean, while the dark grey region is the approximation of the 95% confidence region on the bivariate mean described above. Finally, the light grey region represents, as in the previous examples, the 0.15-expectile region.

The dot marked as a possible outlier at both, the BExPlot and the expectile plot of variable AHF antigen ($y$), corresponds to a non-carrier of Hemophilia A.

# 7 Highlights and conclusions

The main achievement of this paper is the introduction of a new notion of depth, namely the expectile one and the construction of algorithms for its computation on bivariate datasets. The paper starts with a review of the concept of univariate expectile function and a summary of its main properties, while the inverse expectile function is introduced, and some emphasis is placed on their empirical counterparts.

Expectile regions are presented in terms of halfspaces determined by univariate expec-

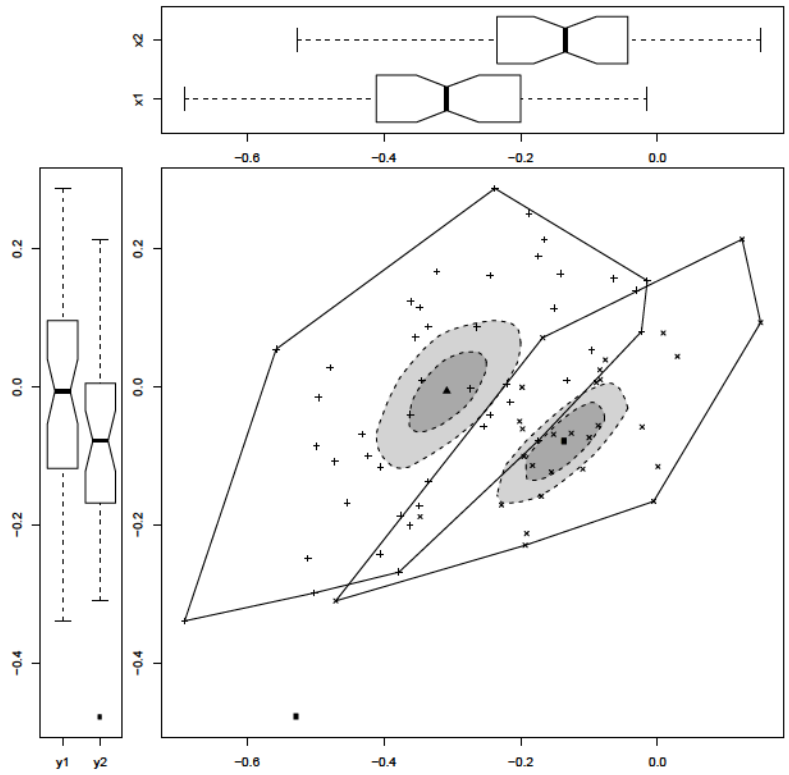



Figure 5: BExPlots with confidence regions (for two populations)

tiles. Due to the fact that expectiles are positively homogeneous and subadditive functions for specific values of $\alpha$, we represent an expectile region as the compact convex set whose support function is the expectile function with respect to a univariate projection of the data. Some basic properties of the expectile regions, as well as other relative to continuity and distribution characterization are discussed. At the same time we present the empirical expectile regions, with an algorithm to determine their extreme points in the bivariate setting, and show their consistency.

Along with the concept of expectile regions, the expectile depth is introduced in a natural way. We discuss the main properties of the expectile depth function and remark the implications of some of those properties on the topology of the expectile regions. We discuss the sample expectile depth function and show its uniform consistency. A second algorithm

to compute the expectile depth of a point w.r.t. a multivariate dataset is presented here.

Finally, we introduce the BExPlot as a practical tool for data visualization and outlier detection in bivariate datasets which can be used to represent the bivariate interactions of higher dimensional data. Unlike the bagplot, which is the most popular bivariate extension of the boxplot, the BExPlot is centred in the mean of a dataset and this allows us to represent an approximation of the confidence region on the mean together with the BExPlot.

## Acknowledgements

We would like to thank Prof. Paul Eilers for presenting his work on expectiles to us and helpful and encouraging comments. This research was partially supported by the Spanish Ministry of Science and Innovation under grant ECO2015-66593-P.

# A Algorithms

## A.1 Algorithm for extreme points of the expectile regions in $\mathbb{R}^2$

We use the classical method of the circular sequence, which provides us with all possible sortings of a bivariate data cloud with respect to their univariate projections in a fast way, to develop an algorithm to compute the extreme points of the *expectile trimmed region* of a bivariate data set. Notice that the first 5 steps in the algorithm below are standard in any circular sequence routine, such as those in Ruts and Rousseeuw (1996); Dyckerhoff (2000); Cascos (2007); Liu *et al.* (2012).

**Basic guidelines of an algorithm to find the extreme points of $\mathrm{E}_n^\alpha$**

*Input:*

- Data points $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$, $i = 1, \ldots, n$.
- Trimming level $0 < \alpha \leq 1/2$.

*Output:*

- Extreme points of $\mathrm{E}_n^\alpha$ supported by half-planes with outer normal $(\cos\beta, \sin\beta)$ with $0 \leq \beta \leq \pi$ (northern boundary).

*Steps:*

Step 1. Store all data points in an $n \times 2$-array and sort them according to the following rule:

$$\boldsymbol{x}_i < \boldsymbol{x}_j \text{ if and only if } (x_{i,1} < x_{j,1}) \text{ or } (x_{i,1} = x_{j,1} \text{ and } x_{i,2} > x_{j,2}).$$

Step 2. Initialize an $n$-array called $\mathbf{R}$ such that $\mathbf{R}_i = i$ for all $i$. Entry $\mathbf{R}_i$ will represent the relative position of point $\boldsymbol{x}_i$ for the ordering given by the one-dimensional projection under consideration (in each iteration of the main loop, Steps 7. to 11.).

Step 3. Compute the angle $\gamma_{i,j}$ of the line defined by each pair of points $\boldsymbol{x}_i, \boldsymbol{x}_j$ and the positive horizontal axis. Sort the angles in an increasing way in a matrix called **ANG** with $\binom{n}{2}$ rows and 3 columns. The $r$-th row of **ANG** contains the $r$-th smallest angle $\gamma_{i,j}$, value $i$, and value $j$. All rows with the same first entry (angle) are ordered in terms of their second entries (x-coordinate), and all rows with the same first and second entries are ordered in terms of their third entries (y-coordinate).

Step 4. Compute the univariate expectile of the x-coordinates of the data, $e_\alpha(x_{1,1}, \ldots, x_{n,1})$, and take $s$ as the sum of those $x_{i,1}$'s that are less than the expectile.

Step 5. Initialize an array called **EXT** to store the extreme points and establish 0 as the first angle to be considered, $\mathbf{ANG}_{0,1} = 0$. Entries $\mathbf{ANG}_{0,2}$ and $\mathbf{ANG}_{0,3}$ are left blank.

Step 6. First iteration of the main loop, set $k = 1$.

Step 7. Compute the candidate to extreme point $\boldsymbol{x}$ for the data sorting given by the array **R** (data points ordered with respect to the univariate projection given by their scalar product times $(\cos \mathbf{ANG}_{k-1,1}, \sin \mathbf{ANG}_{k-1,1})$) and the natural number $s$ as in (11). Observe that the first time we go through this step, the extreme point supported by the halfplane $\{(x, y) : x \leq e_\alpha(x_{1,1}, \ldots, x_{n,1})\}$ is obtained. Every other time a new candidate is computed, only one point is added to the sum in equation (11) when $s$ decreases by one unit, while one point is subtracted when $s$ increases by one unit. If a new data sorting is considered (iteration $k + 1$) only one pair of points interchange their relative positions in the **R** array, and thus, depending on $s$, at most one point is subtracted, while another is added in the sum in Equation (11).

Step 8. Check if the point computed in Step 7. is indeed an extreme point for that $s$ and data sorting. It will be an extreme point as long as its univariate projection through $(\cos \beta, \sin \beta)$ for some $\mathbf{ANG}_{k-1,1} \leq \beta \leq \mathbf{ANG}_{k,1}$ is the expectile of the corresponding univariate projection of the dataset. This will hold as long as the univariate projection

of the candidate through $(\cos\beta, \sin\beta)$ lies between the univariate projections of the $s$-th and the $(s+1)$-th data points in the current data sorting.

Step 9. Consider consecutive values of $s$ and check if there are other extreme points for the same data sorting and such values of $s$. That is, go to Step 7. with $s' = s + 1$ and while the candidate results as extreme in Step 8., try $s' + 1$ in Step 7. Do the same with $s' = s - 1$ and $s' - 1$.

Step 10. Consider the angle between the line through the mean and each extreme point and the x-axis. Append the extreme points in increasing way with regard to those angles in the array **EXT**. Update $s$ as the value that corresponds to the last point in **EXT**.

Step 11. While $k < \binom{n}{2}$, interchange the values of positions $\mathbf{ANG}_{k,2}$ and $\mathbf{ANG}_{k,3}$ of array **R**. Set $k \leftarrow k + 1$ and go to Step 7.

Remember that algorithm A.1 was designed to obtain the extreme points supported by halfplanes with outer normal of the form $(\cos\beta, \sin\beta)$ with $0 \leq \beta \leq \pi$ (the northern boundary). In order to get the remaining points (the southern boundary) transform all data points by reflecting them with respect to the line $y = 0$, apply the algorithm to the transformed data points, and finally reverse the reflection process.

**Complexity**

1. Computing and ordering the $\binom{n}{2}$ angles of all the pairs of points in Step 3. requires $O(n^2 \log n)$ operations.

2. The first computation of a univariate expectile, performed in Step 4., can be done in an exact manner (alternatively to using a Newton-gradient algorithm). It would require sorting all the x-coordinates of the data points, which is done in $O(n \log n)$ operations. If all the possible values of $s$ were to be considered, the complexity would increase at most up to $n^2$.

3. The main loop (Steps 7. to 11.) is run $\binom{n}{2}$ times, but the number of operations at each iteration does not depend on the sample size $n$ because only one point is being added or subtracted. Since $s$ varies in a bounded set of the form $s \pm p$ for some value $p$ (in all of our numerical experiments we found that $0 \leq p \leq 3$), then the complexity remains $O(n^2 \log n)$.

## A.2 Algorithm for expectile depth in $\mathbb{R}^2$

In a similar way to Rousseeuw and Ruts (1996) for the algorithms of the halfspace and simplicial depth, see Liu (1990), the current algorithms proceeds over the projection of the data cloud on a circumference.

*Input:*

- Data points $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$, $i = 1, \ldots, n$.
- Point $\boldsymbol{y}$ for which the expectile depth is computed.

*Output:*

- Depth $\mathrm{ED}_n(\boldsymbol{y})$ of point $\boldsymbol{y}$ with respect to the cloud $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$, $i = 1, \ldots, n$.

*Steps:*

Step 1. Center all data points w.r.t. point $\boldsymbol{y}$.

Step 2. Compute the sample mean $\overline{\boldsymbol{x}}$, find its angle w.r.t. the positive horizontal axis, and rotate the data cloud w.r.t. this angle so that $\overline{\boldsymbol{x}}$ can be written as $\overline{\boldsymbol{x}} = \|\overline{\boldsymbol{x}}\|(0, -1)$ after rotation.

Step 3. For each data point $\boldsymbol{x}_i$ compute the angle $\gamma_i$ between the positive horizontal axis and the ray from the origin to $\boldsymbol{x}_i$.

Step 4. Reflect through the origin those points whose angles are between $\frac{\pi}{2}$ and $\frac{3\pi}{2}$ by subtracting $\pi$ to all the angles, so that the range of angles is $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and the algorithm runs only on a semi-circle. The reflected points are tagged with value $-1$, the rest with value $+1$.

Step 5. Store all data points in an $n \times 4$ array called **ANG** whose entries are the two coordinates of each point, the corresponding angle and the corresponding tag, respectively. Sort array **ANG** with regard to the angles.

Step 6. Set $\gamma_0 = -\pi/2$ and $i = 1$.

Step 7. Compute the average of the points in the halfplane with inner normal $\boldsymbol{u} = (\cos(\gamma_i + \pi/2), \sin(\gamma_i + \pi/2))$, call it $\overline{\boldsymbol{x}}(\gamma_i)$, and compute the angle between $\overline{\boldsymbol{x}}(\gamma_i)$ and the positive horizontal axis, call it $\beta_i$ .

Step 8. Apply the criteria for finding the minimum angle given by the KKT conditions, according to (13), between $\gamma_{i-1}$ and $\gamma_i$ and store this value in an array called **MIN**.

Step 9. While $i < n$, set $i \to i+1$, go to Step 7. and update $\overline{\boldsymbol{x}}(\gamma_i)$ by adding or subtracting the point whose coordinates are $(\mathbf{ANG}_{i,1}, \mathbf{ANG}_{i,2})$, according to the tag it was labeled with in Step 4.

Step 10. With the minimum of those values stored in **MIN** compute the depth according to equation (12) and return it.

**Complexity** The sorting of the $n$ angles in Step 5. has complexity $O(n \log n)$. The main loop (Steps 7. to 9.) is repeated $n$ times, but each updating of $\overline{\boldsymbol{x}}(\gamma_i)$ only involves the addition or subtraction of a point, so its complexity is $O(n)$ and does not affect the overall complexity of algorithm A.2, which remains $O(n \log n)$.

# B Mathematical proofs

## B.1 Proof of Proposition 3.1

Fix $\boldsymbol{u} \in \mathbb{S}^{d-1}$ and consider a sequence $\alpha_n$ such that $\lim_n \alpha_n = \alpha$ for some $0 < \alpha \leq 1/2$, since the expectiles are continuous w.r.t. the parameter $\alpha$, see property 8. in Section 2.1, then we have that

$$\lim_n e_{1-\alpha_n}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle) = e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle)\,.$$

Now, according to Schneider (1993, Th.1.8.12), the pointwise and uniform convergence of support functions on $\mathbb{S}^{d-1}$ are equivalent, thus $\lim_n d_H(\mathrm{E}^{\alpha_n}(\boldsymbol{X}), \mathrm{E}^{\alpha}(\boldsymbol{X})) = 0$. □

## B.2 Proof of Proposition 3.2

Since the expectile regions are nested, we have that if $\alpha > \beta$, then $\mathrm{E}^{\alpha}(\boldsymbol{X}) \subset \mathrm{E}^{\beta}(\boldsymbol{X})$. Assume now that the assertion we want to prove is not valid, that is, there exists $\boldsymbol{y} \in \mathrm{E}^{\alpha}(\boldsymbol{X})$ such that it does not belong to the interior of $\mathrm{E}^{\alpha}(\boldsymbol{X})$, so it lies on its boundary, $\boldsymbol{y} \in \partial \mathrm{E}^{\beta}(\boldsymbol{X})$. As a consequence there is a sequence $\{\boldsymbol{y}_n\}_n$ completely contained in the complement of $\mathrm{E}^{\beta}(\boldsymbol{X})$ such that $\lim_n \boldsymbol{y}_n = \boldsymbol{y}$.

Since each $\boldsymbol{y}_n$ does no lie in $\mathrm{E}^{\beta}(\boldsymbol{X})$ and by the strict parameter monotonicity of expectiles, see property 9. in Section 2.1, it holds

$$\langle \boldsymbol{y}_n, \boldsymbol{u} \rangle > e_{1-\beta}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle) > e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle) \quad \text{for some } \boldsymbol{u} \in \mathbb{S}^{d-1},$$

and therefore for such $\boldsymbol{u}$,

$$\langle \boldsymbol{y}, \boldsymbol{u} \rangle = \lim_n \langle \boldsymbol{y}_n, \boldsymbol{u} \rangle \geq e_{1-\beta}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle) > e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle),$$

which contradicts the fact that $\boldsymbol{y} \in \mathrm{E}^{\alpha}(\boldsymbol{X})$. □

## B.3 Proof of Proposition 4.1

An explicit expression for the expectile depth can be obtained from its definition as a supremum, the formula for the inverse expectile function in (6), and elementary algebra,

$$\begin{aligned}
\mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) &= \sup\{\alpha : \boldsymbol{y} \in \mathrm{E}^{\alpha}(\boldsymbol{X})\} \\
&= \sup\left\{\alpha : \langle \boldsymbol{y}, \boldsymbol{u}\rangle \leq e_{1-\alpha}(\langle X, \boldsymbol{u}\rangle) \quad \text{for all } \boldsymbol{u} \in \mathbb{S}^{d-1}\right\} \\
&= \sup\left\{\alpha : e^{-1}_{\langle \boldsymbol{X}, \boldsymbol{u}\rangle}(\langle \boldsymbol{y}, \boldsymbol{u}\rangle) \leq 1-\alpha \quad \text{for all } \boldsymbol{u} \in \mathbb{S}^{d-1}\right\} \\
&= \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \left(1 - e^{-1}_{\langle \boldsymbol{X}, \boldsymbol{u}\rangle}(\langle \boldsymbol{y}, \boldsymbol{u}\rangle)\right) \\
&= \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \left(1 - \frac{\mathbb{E}\langle \boldsymbol{y}-\boldsymbol{X}, \boldsymbol{u}\rangle + \mathbb{E}\langle \boldsymbol{X}-\boldsymbol{y}, \boldsymbol{u}\rangle_+}{\mathbb{E}\langle \boldsymbol{y}-\boldsymbol{X}, \boldsymbol{u}\rangle + 2\mathbb{E}\langle \boldsymbol{X}-\boldsymbol{y}, \boldsymbol{u}\rangle_+}\right) \\
&= \left(2 + \sup_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\langle \boldsymbol{y}-\mathbb{E}\boldsymbol{X}, \boldsymbol{u}\rangle}{\mathbb{E}\langle \boldsymbol{X}-\boldsymbol{y}, \boldsymbol{u}\rangle_+}\right)^{-1} \\
&= \left(2 - \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\langle \mathbb{E}\boldsymbol{X}-\boldsymbol{y}, \boldsymbol{u}\rangle}{\mathbb{E}\langle \boldsymbol{X}-\boldsymbol{y}, \boldsymbol{u}\rangle_+}\right)^{-1}.
\end{aligned}$$

□

## B.4 Proof of Proposition 4.2

a) Without loss of generality, consider a random vector centred at the origin of coordinates, $\mathbb{E}\boldsymbol{X} = \boldsymbol{0}$, any $\boldsymbol{x} \in \mathbb{R}^d$, and $0 < \lambda < 1$, we just have to show that $\mathrm{ED}(\lambda\boldsymbol{x}; \boldsymbol{X}) > \mathrm{ED}(\boldsymbol{x}; \boldsymbol{X})$.

After the strong projection property, $\alpha = \mathrm{ED}(\boldsymbol{x}; \boldsymbol{X}) = \mathrm{ED}(\langle \boldsymbol{x}, \boldsymbol{u}\rangle, \langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ for some $\boldsymbol{u} \in \mathbb{S}^{d-1}$, now the strict monotonicity of inverse expectile function guarantees that $\mathrm{ED}(\langle \lambda\boldsymbol{x}, \boldsymbol{u}\rangle, \langle \boldsymbol{X}, \boldsymbol{u}\rangle) > \alpha$. Finally, $\mathrm{ED}(\langle \lambda\boldsymbol{x}, \boldsymbol{u}\rangle, \langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is a lower bound of $\mathrm{ED}(\lambda\boldsymbol{x}; \boldsymbol{X})$.

b) As shown in Dyckerhoff (2017, Th.3.2), part b) is equivalent to the continuity of the map $\alpha \mapsto \mathrm{E}^{\alpha}(\boldsymbol{X})$ w.r.t. the Hausdorff metric proved in Proposition 3.1.

□

## B.5 Proof of Theorem 5.1

a) Following Holzmann and Klar (2016, Th.2), for any fixed $\boldsymbol{u} \in \mathbb{S}^{d-1}$ and $\alpha \in (0, 1/2)$, the sample expectiles $e_{1-\alpha}(\langle \boldsymbol{x}_1, \boldsymbol{u} \rangle, \ldots, \langle \boldsymbol{x}_n, \boldsymbol{u} \rangle)$ converge a.s. to $e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle)$. In terms of the support functions of the expectile regions, $\{h(\mathrm{E}_n^\alpha, \boldsymbol{u}) : n \geq 1\}$ converge a.s. to $h(\mathrm{E}^\alpha(\boldsymbol{X}), \boldsymbol{u})$.

According to Molchanov (2017, Prop.1.8.17) the almost sure pointwise (on $\boldsymbol{u}$) convergence of support functions of random compact convex sets to the support function of a deterministic set implies the almost sure convergence of the random sets to the deterministic one in the Hausdorff metric. In conclusion, $\{\mathrm{E}_n^\alpha : n \geq 1\}$ converges a.s. to $\mathrm{E}^\alpha(\boldsymbol{X})$ in the Hausdorff metric.

Following Dyckerhoff (2017, Th.4.7), the strict monotonicity condition of the expectile depth shown in part b) of Proposition 4.2 and the almost sure convergence of $\{\mathrm{E}_n^\alpha : n \geq 1\}$ in the Hausdorff metric for any fixed $\alpha$ imply the almost sure uniform convergence of $\{\mathrm{E}_n^\alpha : n \geq 1\}$ in the Hausdorff metric on a compact set $\alpha \in I \subset (0, 1/2)$. In conclusion,

$$\sup_{\alpha \in I} d_H(\mathrm{E}_n^\alpha, \mathrm{E}^\alpha(\boldsymbol{X})) \xrightarrow{a.s.} 0.$$

b) After property 2. in Section 4.1, the map $\boldsymbol{y} \mapsto \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X})$ is continuous. Together with the result in part a), Dyckerhoff (2017, Th.4.6) has shown that this implies the almost sure uniform convergence of the sequence of depths $\{\mathrm{ED}_n(\,\cdot\,; \boldsymbol{X}) : n \geq 1\}$.

□

# References


Bellini, F., Klar B., Müller, A., and Gianin, E.R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics* **54**, 41–48.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika* **75**, 761–771.

Breckling, J., Kokic, P., and Lübke, O. (2001). A note on multivariate M-quantiles. *Statistics & Probability Letters* **55**, 39–44.

Cascos, I. (2007). The expected convex hull trimmed regions of a sample. *Computational Statistics* **22**, 557–569.

Cascos, I. (2010). Data depth: multivariate statistics and geometry. In W. Kendall and I. Molchanov Eds. *New Perspectives in Stochastic Geometry.* Oxford University Press, pp. 398–423.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *Journal of the London Mathematical Society* **s1-11**, 290–294.

Dyckerhoff, R. (2000). Computing zonoid trimmed regions of bivariate data sets. In J. Bethlehem and P. Heijden Eds. *COMPSTAT 2000. Proceedings in computational statistics.* Physica–Verlag, Heidelberg, pp. 295–300.

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

Dyckerhoff, R. (2017). Convergence of depths and depth-trimmed regions. `https://arxiv.org/abs/1611.08721v2`.

Eilers P. (2010). Expectile contours and data depth. In A.W. Bowman Ed. *Proceedings of the 25th International Workshop on Statistical Modelling.* pp. 167–172.

Giorgi, E. and McNeil, A.J. (2016). On the computation of multivariate scenario sets for the skew-$t$ and generalized hyperbolic families. *Computational Statistics and Data Analysis* **100**, 205–220.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106**, 746–762.

Herrmann, K., Hofert, M., and Mailhot, M. (2018). Multivariate geometric expectiles. *Scandinavian Actuarial Journal* **7**, 629–659.

Holzmann, H. and Klar, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics* **10**, 2355–2371.

Jones, M.C. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters* **20**, 149–153.

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

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

Liu, X., Zuo, Y., and Wang, Z. (2012). Exactly computing bivariate projection depth contours and median. *Computational Statistics and Data Analysis* **60**, 1–11.

Maume-Deschamps, V., Rullière, D., and Said, K. (2016). Multivariate extensions of expectile risk measures. In G. Puccetti Ed. *Dependence Modelling* De Gruyter **5**, 20–44.

Molchanov, I. (2017). Theory of random sets. Springer, London, 2nd edition.

Müller, A. and Stoyan, D. (2002). Comparison Methods for Stochastic Models and Risks. Wiley Series in Probability and Statistics, Wiley, New York.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrika* **55**, 819–847.

Pokotylo, O., Mozharovskyi, P., Dyckerhoff, R., and Nagy, S. (2019). ddalpha: Depth-Based Classification and Calculation of Data Depth. R package version 1.3.9. `https://cran.r-project.org/web/packages/ddalpha/index.html`.

Rousseeuw, P. and Ruts, I. (1996). Bivariate Location Depth. *Computational Statistics and Data Analysis* **23**, 153–168.

Rousseeuw, P., Ruts, I., and Tukey, J. (1999). The Bagplot: A Bivariate Boxplot. *The American Statistician* **53**, 382–387.

Ruts, I. and Rousseeuw, P. (1996). Computing depth contours of bivariate point clouds. *Computational Statistics and Data Analysis* **23**, 153–168.

Schneider, R. (1993). Convex Bodies: The Brunn-Minkowski Theory. Encyclopedia of Mathematics and Its Applications 44, Cambridge University Press, Cambrigde.

Sobotka, F., Schnabel, S., Waltrup, L.S., Eilers, P., and Kauermann, G. (2014). expectreg: Expectile Regression. R package version 0.39. `https://cran.r-project.org/web/packages/expectreg/index.html`.

Tukey, J. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematitians* (Vancouver, 1974) Vol. 2, pp. 523–531.

Wolf, H.P. (2018). aplpack: Another Plot Package 1.3.2. `https://cran.r-project.org/web/packages/aplpack/index.html`.

Ziegel, J. (2016). Coherence and elicitability. *Mathematical Finance* **26**, 901–918.

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