Journal of Multivariate Analysis 184 (2021) 104757

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Expectile depth: Theory and computation for bivariate datasets

Ignacio Cascos *, Maicol Ochoa

Department of Statistics, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Received 16 July 2020
Received in revised form 19 March 2021
Accepted 19 March 2021
Available online 31 March 2021

AMS subject classifications:
62H05
60D05
65C60

Keywords:
Algorithm
Bagplot
Data depth
Depth region
Expectile

ABSTRACT

Expectiles are the solution to an asymmetric least squares minimization problem for univariate data. They resemble the quantiles, and just like them, expectiles are indexed by a level α in the unit interval. In the present paper, we introduce and discuss the main properties of the (multivariate) expectile regions, a nested family of sets, whose instance with level $0 < \alpha \leq 1/2$ is built up by all points whose univariate projections lie between the expectiles of levels α and $1 - \alpha$ of the projected dataset. Such 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 propose here algorithms for determining all the extreme points of the bivariate expectile regions as well as for computing the depth of a point in the plane. We also study the convergence of the sample expectile regions to the population ones and the uniform consistency of the sample expectile depth. Finally, we present some real data examples for which the Bivariate Expectile Plot (BExPlot) is introduced.

1. Introduction

Expectiles were first introduced by Newey and Powell [36] in the context of linear regression as the solution to a minimization problem. They were so named because they resemble the quantiles of a random variable, but unlike them, they are based on a quadratic loss function, as it is the case of the expectation. They have received high attention in several areas such as risk measurement because, under certain conditions, expectiles happen to be elicitable and coherent risk measures, as it was shown by Gneiting [21], Bellini et al. [2], Ziegel [50], and Krätschmer and Zähle [27], and are now capturing attention in data analytic fields such as multiple output M-quantile regression, see [11], or the building of measures of skewness, see [17].

Inspired by Tukey [46], who employed the quantiles of the univariate projections of a bivariate data cloud to produce a family of central (depth) regions, Eilers [18], and later on Giorgi and McNeil [20], suggested to use the expectiles to build the expectile regions of a multivariate dataset as the intersection of the halfspaces whose supporting hyperplanes are determined by the expectiles of univariate projections of the data. Similarly, Daouia and Paindaveine [11] have proposed to use M-quantiles defined as hyperplanes whose independent term is an expectile of a univariate projection of a dataset.

In the past, other multivariate generalizations of the expectiles have been considered by Breckling and Chambers [3], who proposed a class of multivariate M-quantiles as the solution to minimization problems similar to those giving rise to quantiles and expectiles. Since some of those M-quantiles lie out of the convex hull of the dataset they were built from, in a posterior paper Breckling et al. [4] presented an alternative definition of multivariate M-quantiles which still lack to be

* Corresponding author.
E-mail addresses: ignacio.cascos@uc3m.es (I. Cascos), maicol.ochoa@uc3m.es (M. Ochoa).

https://doi.org/10.1016/j.jmva.2021.104757
0047-259X/

equivariant under arbitrary affine transformations. Later on, Maume-Deschamps et al. [31] 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. [22] introduced yet another multivariate version of expectiles, the geometric expectiles, as the unique solution of a convex risk minimization problem. They claim that the homogeneity and equivariance under orthogonal transformations properties of their geometric expectiles make them appealing multivariate risk measures.

The expectile regions, see [11,18,20], are a nested family of sets parametrized by a level, in the sense that the inner regions have a greater level associated with them. The same way that Tukey's central regions inspired the definition of the most popular depth function, the so-called halfspace depth, see [39], the expectile regions induce the expectile depth. We present here a simple description of the expectile depth and its properties, complementary to that of Daouia and Paindaveine [11], and based on the extensive studies of depth functions, their properties, and applications developed by Liu et al. [30], Zuo and Serfling [51], Dyckerhoff [13], or Cascos [6]. As an application, 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. [40].

The main contributions of this manuscript are: (i) the comprehensive study of the properties of the expectile regions, including the Hausdorff consistency of the sample expectile regions; (ii) the introduction of the expectile depth together with the study of its main properties, including the strong uniform consistency of the sample expectile depth; (iii) the construction of algorithms for the exact computation of the expectile regions and depth of bivariate datasets; (iv) the introduction of a new data visualization technique, called the BExPlot. It is further shown that (v) the expectiles and expectile regions characterize distributions with finite first moment, and (vi) some links of the expectile regions with the theory of stochastic orderings are explored.

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 main properties, study the consistency of the sample expectile regions with respect to the Hausdorff metric, and characterize the set of extreme points of the sample expectile regions in a way that will be later on used to describe an algorithm for its computation. In Section 4 we present the expectile depth, study the uniform consistency of the sample expectile depth function, and discuss the computation of its 2-dimensional empirical version. Finally, in Section 5 we introduce the Bivariate Expectile Plot as an EDA tool and some D–D plots built out of expectile depths, while a final discussion highlighting the main contributions of the paper is included in Section 6. Two appendices are placed at the end of the manuscript, Appendix A contains the description of two algorithms, one for the computation of the set of extreme points of the expectile regions of a bivariate dataset, and the other for the computation of the expectile depth, while Appendix B contains the proofs of some mathematical results.

2. Univariate expectiles

Given a random variable X defined on a general probability space and with finite second moment, $\mathrm{E}|X|^2 < \infty$, and given $\alpha \in (0, 1)$, the α-expectile of X was defined in Newey and Powell [36] as the minimizer of the quadratic expression

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

where $a_+ = \max\{a, 0\}$ and $a_- = -\min\{a, 0\}$ for any $a \in \mathbb{R}$. Observe that the α-quantile of X is the minimizer of an expression similar to (1), but built without 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

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

An immediate consequence of (2) is that the expectiles are computable for random variables with the only condition of having finite first moment.

Alternative expressions for the expectiles are presented below. Their in detail derivation from (2) is shown in Appendix B.1, and they will be of use when interpreting and computing expectiles. It is clear from them that the α-expectile of a random variable X is a weighted average of the gravity centres of the lower and upper tail of X with regard to itself such that the weight of the lower gravity centre corresponds to the weight of the upper gravity centre in a proportion of $(1-\alpha)$ to α. Specifically it holds

$$e_\alpha(X) = \mathrm{E}X + \frac{2\alpha-1}{1-\alpha}\mathrm{E}(X-e_\alpha(X))_+ = \frac{(1-\alpha)\int_0^{F_X(e_\alpha(X))} F_X^{-1}(t)\,\mathrm{d}t + \alpha\int_{F_X(e_\alpha(X))}^1 F_X^{-1}(t)\,\mathrm{d}t}{\alpha + (1-2\alpha)F_X(e_\alpha(X))}\,, \tag{3}$$

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

Let us describe next a geometric interpretation of the expectiles. In the light of the Choquet integral of random variables $(X-e_\alpha(X))_-$ and $(X-e_\alpha(X))_+$, it follows from (2) that the α-expectile of X is the value such that the area between the

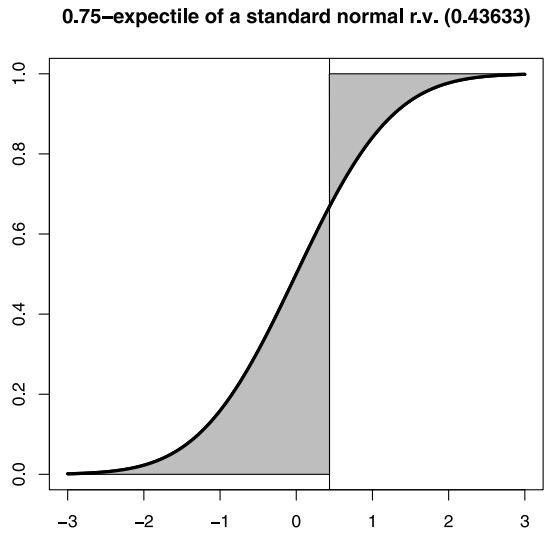


Fig. 1. The thick solid line represents the cdf of a standard normal random variable and the vertical line is established at its 0.75-expectile, $x = 0.43633$. The area of the shaded region to the left of the vertical line is $0.75/(1-0.75) = 3$ times the area of the shaded region to its right.

graph of the 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,

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

See [9] for the Choquet integral and Fig. 1 for a graphical representation of this geometric interpretation.

2.1. Properties of the expectiles and the inverse expectile function

We list below the main properties of the univariate expectiles, whose proofs can be found in classical references, specifically Bellini et al. [2]. As presented in properties (i) to (x), the expectile of X of level $\alpha = 1/2$ is the mean of X, the expectile of X of level α is the negative of the expectile of $-X$ of level $1-\alpha$, expectiles are equivariant with respect to deterministic translations, positive homogeneous, and strictly monotone in their random argument. Further, the expectiles of levels $\alpha \geq 1/2$ are subadditive, while those of level $\alpha \leq 1/2$ are superadditive. Finally, expectiles are continuous and strictly increasing in the parameter α (if X is a nondegenerate random variable).

(i) *Most central expectile:* $e_{1/2}(X) = \mathrm{E}X$;
(ii) *Upper and lower expectiles:* $e_\alpha(X) = -e_{1-\alpha}(-X)$;
(iii) *Translation equivariance:* $e_\alpha(X+a) = a + e_\alpha(X)$, $a \in \mathbb{R}$;
(iv) *Homogeneity:* $e_\alpha(\lambda X) = \lambda e_\alpha(X)$, $\lambda \geq 0$;
(v) *Monotonicity:* if $X \leq Y$ a.s., then $e_\alpha(X) \leq e_\alpha(Y)$;
(vi) *Strict monotonicity:* if $X \leq Y$ a.s. and $\Pr(X < Y) > 0$, then $e_\alpha(X) < e_\alpha(Y)$;
(vii) *Subadditivity:* for $1/2 \leq \alpha < 1$, $e_\alpha(X+Y) \leq e_\alpha(X) + e_\alpha(Y)$;
(viii) *Superadditivity:* for $0 < \alpha \leq 1/2$, $e_\alpha(X+Y) \geq e_\alpha(X) + e_\alpha(Y)$;
(ix) *Parameter continuity:* e_α is continuous in α;
(x) *Strict parameter monotonicity:* e_α is strictly increasing in parameter α whenever random variable X is not degenerate.

The inverse of the expectile function plays a relevant role when showing that expectiles characterize distributions. While for any fixed random variable, the expectile function, whose argument is the level α, plays a similar role to the quantile function, Jones [24] discusses the inverse expectile function as an analogue to the cumulative distribution function (inverse quantile). This function will become relevant when we study the expectile depth in Section 4.

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

$$e_X^{-1}(x) = \left(1 + \frac{\mathrm{E}(X-x)_+}{\mathrm{E}(X-x)_-}\right)^{-1} = \frac{\mathrm{E}(X-x)_-}{\mathrm{E}|X-x|} = \frac{x - \mathrm{E}X + \mathrm{E}(X-x)_+}{x - \mathrm{E}X + 2\mathrm{E}(X-x)_+}. \tag{4}$$

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

The expectile function of a random variable (with finite first moment) characterizes its distribution. This can be verified by obtaining an explicit expression of the stop-loss function, see [34], associated with X out of (4),

$$\mathrm{E}(X-x)_+ = \frac{(x-\mathrm{E}X)(1-e_X^{-1}(x))}{2e_X^{-1}(x)-1}. \tag{5}$$

The expectile function determines the inverse expectile function, together with the mean of a random variable (expectile of level 1/2), and consequently its stop-loss function, see (5). Finally, after the relation $\mathrm{E}(X-x)_+ = \int_x^\infty (1-F_X(t))\,\mathrm{d}t$, the stop-loss function (sometimes called integrated survival function) characterizes the distribution of a random variable.

Stochastic orders are partial order relations (or more generally preoders) between distributions of random elements, see [35,44] for two comprehensive introductions to them. Bellini et al. [1] introduced the expectile stochastic ordering for random variables. Specifically, given X, Y two random variables with finite first moment, they say that X is smaller than Y in the expectile order if $e_\alpha(X) \le e_\alpha(Y)$ for every $0 < \alpha < 1$. Such relation is equivalent to the inverse expectile function of Y being not greater than the inverse expectile function of X, that is, $e_Y^{-1}(x) \le e_X^{-1}(x)$ for every $\max\{\operatorname{ess\,inf}(X), \operatorname{ess\,inf}(Y)\} < x < \max\{\operatorname{ess\,sup}(X), \operatorname{ess\,sup}(Y)\}$. After (4), we obtain the equivalent expression

$$\mathrm{E}(X-x)_+(x-\mathrm{E}Y) \le \mathrm{E}(Y-x)_+(x-\mathrm{E}X)\,, \tag{6}$$

where the restriction on x can be omitted. See [1, Th. 8c] for an alternative derivation of (6). Arguing in terms of the expectile function, we can show that a condition of the type $e_\alpha(X) \le e_\alpha(Y)$ for every $0 < \alpha \le \beta < 1$ holds whenever (6) is satisfied for every $x < e_\beta(X)$, while after Property (ii) in Section 2.1, $e_\alpha(X) \ge e_\alpha(Y)$ for every $0 < \beta \le \alpha < 1$ holds whenever the reverse inequality to (6) is satisfied for every $x > e_\beta(X)$. That is, we have

$$\begin{aligned} &e_\alpha(X) \le e_\alpha(Y) \text{ for every } 0 < \alpha \le \beta < 1 \text{ if and only if } \mathrm{E}(X-x)_+(x-\mathrm{E}Y) \le \mathrm{E}(Y-x)_+(x-\mathrm{E}X) \text{ for every } x < e_\beta(X)\,,\\ &e_\alpha(X) \ge e_\alpha(Y) \text{ for every } 0 < \beta \le \alpha < 1 \text{ if and only if } \mathrm{E}(X-x)_+(x-\mathrm{E}Y) \ge \mathrm{E}(Y-x)_+(x-\mathrm{E}X) \text{ for every } x > e_\beta(X)\,. \end{aligned} \tag{7}$$

2.2. Sample expectiles

Consider a sample of univariate observations $x_1, x_2, \ldots, x_n \in \mathbb{R}$ and $0 < \alpha < 1$. The sample α-expectile, denoted by $e_\alpha(x_1, \ldots, x_n)$ or shortly $e_{n,\alpha}$, is the solution to Eq. (3) for an empirical distribution, and results to be a weighted average of the sample observations. The empirical counterparts of (3) adopt the expressions

$$e_{n,\alpha} = \bar{x} + \frac{2\alpha-1}{n(1-\alpha)} \sum_{x_i > e_{n,\alpha}} (x_i - e_{n,\alpha}) = \frac{1-\alpha}{\alpha n + nF_n(e_{n,\alpha})(1-2\alpha)} \left(n\bar{x} + \frac{2\alpha-1}{1-\alpha} \sum_{x_i > e_{n,\alpha}} x_i \right), \tag{8}$$

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

Fortunately enough, $e_{n,\alpha}$ can be computed in a fast way by means of a repeated weighted averaging algorithm, implemented as a built-in function in the R package `expectreg` by Sobotka et al. [45] which is available on GitHub.

For the previous univariate sample and any $\min x_i \le x \le \max x_i$, according to (4), x is the sample expectile at the level given by the empirical inverse expectile function

$$e_n^{-1}(x) = \frac{nx - n\bar{x} + \sum_{x_i > x}(x_i - x)}{nx - n\bar{x} + 2\sum_{x_i > x}(x_i - x)}\,. \tag{9}$$

3. Expectile regions

Following Eilers [18] and the scenario set construction of Giorgi and McNeil [20], for any d-dimensional random vector $\boldsymbol{X}$ with finite first moment $\mathrm{E}\|\boldsymbol{X}\| < \infty$, its expectile region of level $0 < \alpha \le 1/2$, denoted by $\mathrm{ED}^\alpha(\boldsymbol{X})$, is the intersection of closed halfspaces supported by hyperplanes whose constant term is determined by a univariate expectile

$$\mathrm{ED}^\alpha(\boldsymbol{X}) = \bigcap_{\boldsymbol{u} \in \mathbb{S}^{d-1}} \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{u} \rangle \le e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle)\}\,, \tag{10}$$

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 $\mathbb{R}^d$.

Observe that the expectile regions adopt the expressions proposed by Dyckerhoff [13, Th. 5] when considering the interval between the α and $1-\alpha$ expectiles or the M-quantile regions of Daouia and Paindaveine [11, Def. 3.2] for the appropriate loss function.

In the univariate setting, for $d = 1$, the expectile region of level α is the closed interval of real numbers

$$\mathrm{ED}^\alpha(X) = [e_\alpha(X), e_{1-\alpha}(X)]\,. \tag{11}$$

As an intersection of closed halfspaces, the expectile regions are closed convex sets, and thus characterized by their support functions, see e.g. Schneider [43]. The support function of a closed convex non-empty set $K \subseteq \mathbb{R}^d$ 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 positively homogeneous and subadditive, while the reverse does also hold as can be found in [43, 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 $K \subseteq \mathbb{R}^d$ such that h is its support function.

From the homogeneity and subadditivity of univariate expectiles (Properties (iv) and (vii) 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 function of a compact and convex set, which is in fact $\mathrm{ED}^\alpha(\boldsymbol{X})$. Hence, alternatively to (10), the expectile regions can be characterized in terms of its support function as

$$h(\mathrm{ED}^\alpha(\boldsymbol{X}), \boldsymbol{u}) = \sup\{\langle \boldsymbol{x}, \boldsymbol{u}\rangle : \boldsymbol{x} \in \mathrm{ED}^\alpha(\boldsymbol{X})\} = e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle). \tag{12}$$

An important consequence of (12) is that, since $e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$ is the upper end-point of the (univariate) expectile region of $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ of level α, the expectile depth satisfies the strong (also weak) projection property, see Dyckerhoff [13, Th. 3].

3.1. Properties of the expectile regions

The expectile regions fulfill the usual properties for depth regions considered by Dyckerhoff [13] and Cascos [6], that is, if $0 < \alpha \leq 1/2$ then, as shown in Appendix B.2:

(i) *Most central point:* $\mathrm{ED}^{1/2}(\boldsymbol{X}) = \{\mathrm{E}\boldsymbol{X}\}$;
(ii) *Nesting:* $\mathrm{ED}^\alpha(\boldsymbol{X}) \subseteq \mathrm{ED}^\beta(\boldsymbol{X})$, $0 < \beta \leq \alpha \leq 1/2$;
(iii) *Convexity:* $\mathrm{ED}^\alpha(\boldsymbol{X})$ is convex;
(iv) *Compactness:* $\mathrm{ED}^\alpha(\boldsymbol{X})$ is compact;
(v) *Affine equivariance:* $\mathrm{ED}^\alpha(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\mathrm{ED}^\alpha(\boldsymbol{X}) + \boldsymbol{b}$ for any matrix $\boldsymbol{A} \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$;
(vi) *Monotonicity:* if $\boldsymbol{X} \preceq \boldsymbol{Y}$ (componentwisely) a.s., then $\mathrm{ED}^\alpha(\boldsymbol{Y}) \oplus \mathbb{R}^d_+ \subseteq \mathrm{ED}^\alpha(\boldsymbol{X}) \oplus \mathbb{R}^d_+$; 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 $K_1 \oplus K_2 = \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in K_1, \boldsymbol{y} \in K_2\}$.
(vii) *Minkowski subadditivity:* $\mathrm{ED}^\alpha(\boldsymbol{X} + \boldsymbol{Y}) \subseteq \mathrm{ED}^\alpha(\boldsymbol{X}) \oplus \mathrm{ED}^\alpha(\boldsymbol{Y})$.

Notice that the monotonicity and Minkowski subadditivity, which are not among the classical requirements for families of central regions, are particularly relevant when they are used to assess the risk of multi-asset portfolios, as stressed by Cascos and Molchanov [8].

Similarly to the expectile function, the family of expectile regions characterizes the distribution of a random vector with finite first moment. For any d-dimensional random vector $\boldsymbol{X}$ with finite first moment, $\boldsymbol{u} \in \mathbb{R}^d$, and $0 < \alpha \leq 1/2$, the support function of $\mathrm{ED}^\alpha(\boldsymbol{X})$ evaluated on $\boldsymbol{u}$ is the $(1-\alpha)$-expectile of the linear combination of its components $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$, see (12), while $\mathrm{ED}^\alpha(-\boldsymbol{X})$ is $\mathrm{ED}^\alpha(\boldsymbol{X})$ reflected (by the affine equivariance of the expectile regions, Property (v) in Section 3.1) and its support function evaluated on $\boldsymbol{u}$ is the negative of the α-expectile of $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ (Property (ii) in Section 2.1). After the characterization property of the expectiles, the distribution of every $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ is thus determined by $\{\mathrm{ED}^\alpha(\boldsymbol{X})\}_{0<\alpha\leq 1/2}$. Finally, and according to Cramér and Wold [10], the family of all expectile regions of $\boldsymbol{X}$ characterizes its distribution.

In order to discuss the continuity of the expectile regions with respect to the level α, we need to introduce a notion of distance for compact and convex sets. If $K_1, K_2 \subseteq \mathbb{R}^d$ 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 [32, 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 1. *If* $\lim_n \alpha_n = \alpha$ *in* $(0, 1/2]$, *then*

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

See Appendix B.3 for the proof of Proposition 1.

Proposition 2 will be needed in order to show the continuity of the expectile depth, which constitutes the main topic in Section 4.

Proposition 2. *If* $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ *is a nondegenerate random variable for every* $\boldsymbol{u} \in \mathbb{S}^{d-1}$, *then for every* $0 < \beta < \alpha \leq 1/2$

$$\mathrm{ED}^\alpha(\boldsymbol{X}) \subseteq \operatorname{int} \mathrm{ED}^\beta(\boldsymbol{X}).$$

See Appendix B.4 for the proof of Proposition 2.

The inclusion relation between all expectile regions of two random variables of the same level characterizes the convex stochastic order, while the one of two random vectors characterizes the linear convex stochastic order.

Given two random variables X, Y with finite first moment, condition $\mathrm{ED}^\alpha(X) \subseteq \mathrm{ED}^\alpha(Y)$ for every $0 < \alpha \leq 1/2$ holds if $e_\alpha(Y) \leq e_\alpha(X)$ for every $0 < \alpha \leq 1/2$ and $e_\alpha(X) \leq e_\alpha(Y)$ for every $1/2 \leq \alpha < 1$, see (11), which in particular (for $\alpha = 1/2$) imply $\mathrm{E}X = \mathrm{E}Y$. After (7), the stop-loss functions must be ordered as $\mathrm{E}(X - x)_+ \leq \mathrm{E}(Y - x)_+$ for every $x \in \mathbb{R}$ so, after [35, Th. 1.5.7], X is smaller than Y in the increasing convex order. Further, since $\mathrm{E}X = \mathrm{E}Y$, we have that X is smaller than Y in the convex order, see [35, Th. 1.5.3].

Consider now two random vectors $\boldsymbol{X}, \boldsymbol{Y}$ with finite first moment, such that $\mathrm{ED}^\alpha(\boldsymbol{X}) \subseteq \mathrm{ED}^\alpha(\boldsymbol{Y})$ for every $0 < \alpha \leq 1/2$, then, for any $\boldsymbol{u} \in \mathbb{R}^d$, $e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) = h(\mathrm{ED}^\alpha(\boldsymbol{X}), \boldsymbol{u}) \leq h(\mathrm{ED}^\alpha(\boldsymbol{Y}), \boldsymbol{u}) = e_{1-\alpha}(\langle \boldsymbol{Y}, \boldsymbol{u}\rangle)$ and $-e_\alpha(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) = e_{1-\alpha}(\langle \boldsymbol{X}, -\boldsymbol{u}\rangle) = h(\mathrm{ED}^\alpha(\boldsymbol{X}), -\boldsymbol{u}) \leq h(\mathrm{ED}^\alpha(\boldsymbol{X}), -\boldsymbol{u}) = e_{1-\alpha}(\langle \boldsymbol{Y}, \boldsymbol{u}\rangle) = -e_\alpha(\langle \boldsymbol{Y}, \boldsymbol{u}\rangle)$, so $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ is smaller than $\langle \boldsymbol{Y}, \boldsymbol{u}\rangle$ in the convex stochastic order. Since the order relation is satisfied for every $\boldsymbol{u} \in \mathbb{R}^d$, after [35, Def. 3.5.1], we conclude that $\boldsymbol{X}$ is smaller than $\boldsymbol{Y}$ in the linear convex stochastic order. The family of zonoid depth regions introduced by Koshevoy and Mosler [25] does also characterize the linear convex order by means of inclusions, see [26,33].

3.2. Sample expectile 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 α-expectile region is the set ED_n^α whose support function, see (12), is

$$h(\mathrm{ED}_n^\alpha, \boldsymbol{u}) = e_{n,1-\alpha}(\langle \boldsymbol{x}_1, \boldsymbol{u}\rangle, \ldots, \langle \boldsymbol{x}_n, \boldsymbol{u}\rangle)\,, \tag{13}$$

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

Remark 1. Observe that despite the support function of an empirical expectile region evaluated on any $\boldsymbol{u} \in \mathbb{S}^{d-1}$ is an empirical expectile of the projected sample, and after (8), empirical expectiles are weighted averages (in this case of the projected data), the (empirical) expectile regions are not specific instances of the weighted mean trimmed regions proposed by Dyckerhoff and Mosler [15]. This is because the weights in the formula of the support function of the weighted mean trimmed regions depend only on the rank of the projection of each specific data point, while expectiles are the minimizers of a quadratic problem, and the weight that they award to each observation does also depend on the distance to the remaining observations. An immediate consequence is that neither the (population) expectile regions belong to the family of population weighted mean trimmed regions described in [16].

Our first result on the sample expectile regions is their consistency.

Proposition 3. *Given any d-dimensional random vector $\boldsymbol{X}$ with finite first moment, the family of sample expectile regions $\{\mathrm{ED}_n^\alpha\}_\alpha$ built from a random sample of $\boldsymbol{X}$ satisfies*

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

for any compact set $I \subseteq (0, 1/2)$, where d_H stands for the Hausdorff metric.

See Appendix B.5 for the proof of Proposition 3.

In order to compute all the extreme points of the expectile regions, we will first characterize them.

Proposition 4. *The sample α-expectile region can be written as the convex hull of all linear combinations of the points from the data sample with some prescribed weights as*

$$\mathrm{ED}_n^\alpha = \mathrm{co}\left\{ \boldsymbol{x} = \frac{\alpha}{(1-\alpha)n + s(2\alpha - 1)} \left(n\bar{\boldsymbol{x}} + \frac{1-2\alpha}{\alpha} \sum_{\pi_{\boldsymbol{u}}(i) > s} \boldsymbol{x}_i \right) : \boldsymbol{u} \in \mathbb{S}^{d-1} \text{ and } G_{\boldsymbol{u}}(\boldsymbol{x}) = s \right\}, \tag{14}$$

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

The proof of Proposition 4 follows directly applying (8) to all possible sortings of the dataset. Most of the points computed in (14) are not extreme ones, but inner points of ED_n^α instead.

Proposition 4 contains the description of all possible extreme points of the expectile regions, but a brute force algorithm based on it would have complexity $O(n!n^2)$, where $n!$ is the number of permutations of the data points, and it is multiplied times n because of the position that the expectile might occupy, and again times n because of the sums involved in the computation of the extreme points. Instead, in Appendix A.1, we use a circular sequence routine to obtain all extreme points with complexity $O(n^2 \log n)$. Specifically, the algorithm moves efficiently through the $\binom{n}{2}$ sortings of the univariate projections of the data points. For the initial order (x-coordinates in ascending order), we compute the expectile of the first coordinates and the position it occupies inside the sample (of x-coordinates). Then we apply (14) to obtain the first extreme point and check whether it is possible to obtain another extreme point for the same sorting (which corresponds to projections determined by a $\boldsymbol{u}$ lying in an arc of the circumference $\mathbb{S}^1$) but such that it occupies another position in the sorted projected sample. Once we have all extreme points for that given sorting, we consider the next possible sorting of

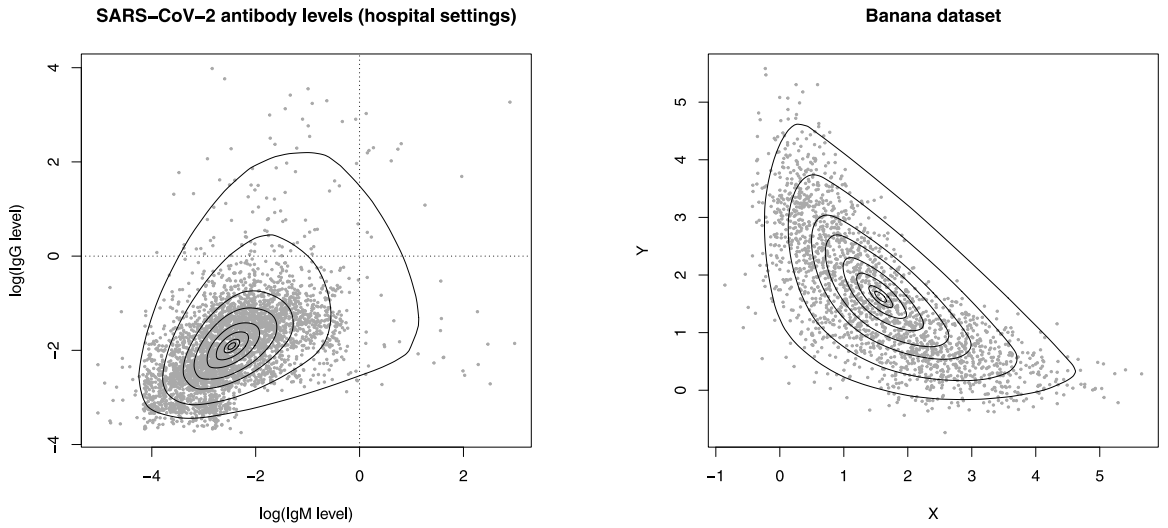


Fig. 2. Contours of the expectile regions of levels 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.45 of two datasets.

the data points in the circular sequence routine. This is done for all possible sortings and positions. Each time we move from one possible extreme point to the next, the number of operations we do is at most two (the position of two points is interchanged) in order keep the computational complexity as low as possible. Notice that the first steps of the algorithm in Appendix A.1 are the standard ones in any circular sequence routine, such as those in [41] for the halfspace regions, [12] for the zonoid regions, and [5] for the expected convex hull regions, all of them with complexity $O(n^2 \log n)$.

As an illustration of the expectile regions introduced in this section, Fig. 2 shows the contours of some expectile regions of two datasets. On the left, we present a scatterplot of the level of antibodies for SARS-CoV-2 built for data taken from patients and staff of Chinese hospitals available at [49]. Axis Y corresponds to the logarithm of the level of IgG antibodies, while axis X corresponds to the logarithm of the level of IgM antibodies. Each antibody level is computed as the ratio of the chemiluminescence signal over the cutoff value, and thus the dotted horizontal and vertical lines represent the cutoff values of antibody levels. The positive correlation between the variables is captured by the expectile regions, which also capture the enlarged variability of each of the variables, when the other one assumes high values.

In Fig. 2 right, we present the scatterplot of a simulated banana-shaped dataset together with some of its expectile regions. Specifically, 2000 observations were simulated with regard to the bivariate distribution with conditional normal distributions described by Gelman and Meng [19] (with parameters $A = 0.5$, $B = 0$, $C_1 = C_2 = 3$). Unlike the level sets of the joint density and the support of the empirical distribution which are banana-shaped, the expectile regions are convex. Nonetheless, these regions contain valuable information about the data cloud. In fact they characterize it.

These regions could be used in statistical process control for the monitorization of the location parameter of multivariate processes taking the sample mean as monitoring statistic. As in [7], the expectile region of some given level would be interpreted as the in-control region for a control chart built for the sample mean, which we recall serves as monitoring statistic. Due to the convexifying effect of averaging, the expectile regions fit reasonably well to the distribution of the sample mean. Clearly, evaluating the centrality of each observation in terms of the level of the expectile region it belongs to, as suggested by Liu [29] on her seminal work on control charts based on notions of data depth might not be the best option here, since the regions do not truly capture the shape of the distribution of the data, but a convexified version of it instead.

4. Expectile depth function

Following the classical introduction of depth functions from central regions [see 5,12,13,41,46,51] and alike [11, Def. 4.1], we propose the expectile depth as the degree of centrality of a point $\boldsymbol{y} \in \mathbb{R}^d$ with respect to the distribution of a d-dimensional random vector $\boldsymbol{X}$ in terms of its expectile regions. Specifically, the expectile depth of $\boldsymbol{y}$ with respect to $\boldsymbol{X}$ is

$$\mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) = \sup\{0 < \alpha \leq 1/2 : \boldsymbol{y} \in \mathrm{ED}^{\alpha}(\boldsymbol{X})\}\,. \tag{15}$$

This way, the expectile region of level α can be rewritten as the set of all the points whose depth is at least α, $\mathrm{ED}^{\alpha}(\boldsymbol{X}) = \{\boldsymbol{y} : \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) \geq \alpha\}$, and thanks to 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 with respect to 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 5 will turn out to be crucial when explicitly computing empirical expectile depths.

Proposition 5. *For any $\boldsymbol{y} \in \mathbb{R}^d$, the expectile depth function satisfies*

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

See Appendix B.6 for the proof of Proposition 5.

The sample expectile depth function is the empirical counterpart of (15), which is obtained by replacing the expectile regions of random vector $\boldsymbol{X}$ by sample expectile regions. Given any point $\boldsymbol{y} \in \mathbb{R}^d$ and a sample $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \subseteq \mathbb{R}^d$, it is clear from Proposition 5 that the sample expectile depth is

$$\mathrm{ED}_n(\boldsymbol{y}) = \left(2 - \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\sum\langle \boldsymbol{x}_i - \boldsymbol{y}, \boldsymbol{u}\rangle}{\sum\langle \boldsymbol{x}_i - \boldsymbol{y}, \boldsymbol{u}\rangle_+}\right)^{-1}. \tag{16}$$

The sample expectile depth is a uniformly consistent estimator of the population expectile depth.

Proposition 6. *Given any d-dimensional random vector $\boldsymbol{X}$ with finite first moment and such that $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ is a nondegenerate random variable for every $\boldsymbol{u} \in \mathbb{S}^{d-1}$, the sample expectile depth function ED_n satisfies:*

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

See Appendix B.7 for the proof of Proposition 6.

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:

(i) *Affine invariance,* the expectile depth is independent of the coordinate system. For any matrix $A \in \mathbb{R}^{d\times d}$ and $\boldsymbol{b} \in \mathbb{R}^d$

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

(ii) *Continuity,* if $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ is a nondegenerate random variable for every $\boldsymbol{u} \in \mathbb{S}^{d-1}$, the mapping $\boldsymbol{y} \mapsto \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X})$ is continuous by Proposition 2 and Dyckerhoff [14, Th. 3.1];

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

(iv) *Quasiconcavity,* as a consequence of the convexity of the expectile regions,

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

(v) *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 [14].

Proposition 7. *Strict monotonicity of the expectile depth.*

(i) The expectile depth is strictly monotone in rays from the centre of the distribution.
(ii) For $0 < \alpha < 1/2$ it holds $\mathrm{ED}^\alpha(\boldsymbol{X}) = \mathrm{cl}\{\boldsymbol{y} : \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X}) > \alpha\}$.

See Appendix B.8 for the proof of Proposition 7.

4.2. Computation of the bivariate expectile depth

Our goal now is to compute the empirical expectile depth function with respect to the sample $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \subseteq \mathbb{R}^2$. For the sake of simplicity, we assume that the point $\boldsymbol{y}$ whose depth is to be assessed is the origin (otherwise, the depth can be computed after centring the sample at the given point). The sample expectile depth of the origin is

$$\mathrm{ED}_n(\boldsymbol{0}) = \left(2 - \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\sum\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle}{\sum\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle_+}\right)^{-1}. \tag{17}$$

We assume further that the sample average, $\bar{\boldsymbol{x}}$, lies on the negative part of the vertical axis (rotate the sample if needed) and write $\boldsymbol{u} \in \mathbb{S}^1$ as $\boldsymbol{u} = (\cos\gamma, \sin\gamma)$ for some $\gamma \in [0, 2\pi)$.

The infimum in $\mathrm{ED}_n(\mathbf{0})$ turns into

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

where γ_i is the angle between the positive horizontal semiaxis and the ray from the origin containing $\boldsymbol{x}_i$, and $\bar{\boldsymbol{x}}(\gamma) = \|\bar{\boldsymbol{x}}(\gamma)\|(\cos\beta, \sin\beta)$ is the sum of the sample points in the halfspace with inner normal $(\cos(\gamma + \pi/2), \sin(\gamma + \pi/2))$ divided by n.

The function to minimize over the interval $(\gamma_i + \pi/2, \gamma_{i+1} + \pi/2)$ is $f(\gamma) = -\sin\gamma/\cos(\gamma - \beta)$, whose monotonicity is explained by the sign of its derivative $f'(\gamma) = -\cos\beta/\cos^2(\gamma - \beta)$, which is positive or negative depending on β, and we have:

$$\begin{aligned}&\text{if } \frac{\pi}{2} < \beta < \frac{3\pi}{2}, \text{ the minimizer is } \gamma^* = \gamma_i + \frac{\pi}{2}; \\ &\text{if } -\frac{\pi}{2} < \beta < \frac{\pi}{2}, \text{ the minimizer is } \gamma^* = \gamma_{i+1} + \frac{\pi}{2}.\end{aligned} \tag{18}$$

In Appendix A.2, we describe an algorithm to compute the bivariate expectile depth that takes advantage of the fact that the angle γ for which $\boldsymbol{u} = (\cos\gamma, \sin\gamma)$ is the minimizer of expression (17) is $\pi/2$ radians distant from the angle that forms the positive horizontal semiaxis and the ray from the origin containing a data point. All such angles are considered and the minimum of the ratio of the sums is transformed as suggested in (17) to obtain an expectile depth. Consequently, and in a similar way to [38] for the algorithms of the halfspace and simplicial depths, see [28], the algorithm described in Appendix A.2 proceeds over the projection of the data cloud on a circumference centred on the point whose depth is evaluated attaining a computational complexity equal to $O(n \log n)$.

5. The Bivariate Expectile Plot (BExPlot), E–E plots, and D–D plots

The α-expectile of a univariate standard normal random variable Z can be computed after (3) and the explicit expression of its stop-loss function. It turns out to be the unique solution to the equation

$$\frac{\phi(e_\alpha(Z))}{e_\alpha(Z)} + \Phi(e_\alpha(Z)) = \frac{\alpha}{2\alpha - 1}, \ \alpha \neq 1/2, \tag{19}$$

where ϕ is the standard normal density and Φ the standard normal cdf, while $e_{0.5}(Z) = 0$. See Fig. 3 for the representation of the expectile function, together with the quantile function of a standard normal random variable. Observe that the 0.15-expectile of a standard normal random variable approximately matches its 0.25-quantile, and the same occurs with the 0.85-expectile and the 0.75-quantile. From the equivariance properties of the expectiles, the former assertions can be generalized to any normal random variable.

5.1. The BExPlot and the eboxplot

We describe next the BExPlot, a bivariate data visualization tool based on the expectile regions. It was inspired on the Bagplot of Rousseeuw et al. [40], so we start with a brief description of it.

The Bagplot is a bivariate generalization of the classical quantile-boxplot built from Tukey's halfspace depth. Its centremost point is the barycentre of the deepest region (with regard to the halfspace depth). A bag consisting in the smallest halfspace region containing 50% of the observations is plotted. This bag is dilated from the centremost point by a factor of 3 in order to obtain an unplotted fence. Points out of the fence are considered as outliers, while a loop is plotted as the contour of the convex hull of all points inside the fence.

Given a univariate dataset, the expectile-boxplot (eboxplot) is similar to the classical quantile-boxplot (qboxplot), see top and right margins of Figs. 4 and 5. It represents a box ranging between the 0.15-expectile and the 0.85-expectile with a mark on the average value (0.5-expectile). Possible outliers to either side of the box are identified as those points located at a distance from the nearest end-point of the box farther than 3 times the distance between the 0.5-expectile and such end-point. They are highlighted with a bullet, while whiskers are represented at the largest and smallest observations excluding those highlighted as possible outliers. The reason to take the 0.15- and 0.85-expectiles is that, as explained above, for a normal distribution, they approximately coincide with its first and third quartiles, see Fig. 3. Consequently, the eboxplot looks very similar to a qboxplot when the underlying distribution is normal.

The Bivariate Expectile Plot (BExPlot) is a graphical representation for bivariate data that consists of a bullet located at the sample mean (singleton representing the 0.5-expectile region), a shaded bag representing the 0.15-expectile region, and a convex barrier 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 barrier (if any) are marked as possible outliers.

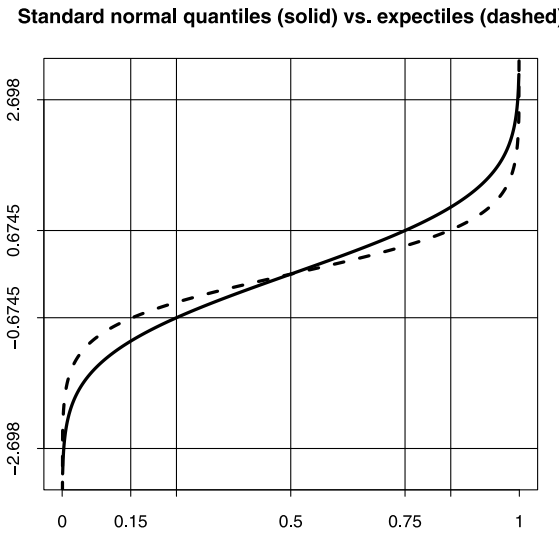


Fig. 3. Quantile function of a standard normal random variable, represented as a solid line, and expectile function of a standard normal random variable, represented as a dashed line.

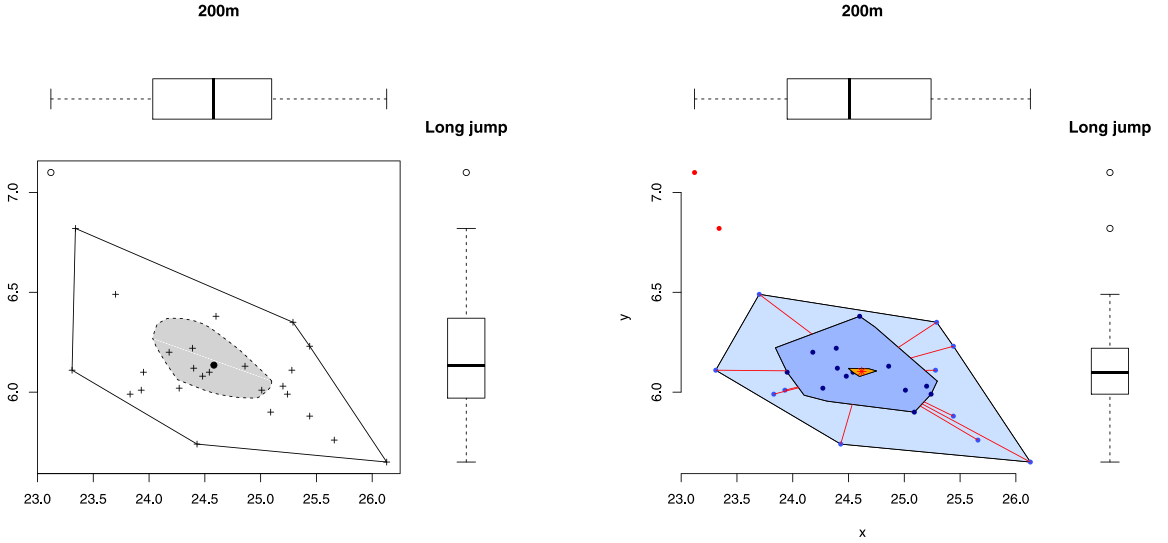


Fig. 4. BExPlot and Bagplot for the 200 metres and long jump of the (women's) Heptathlon in the Barcelona 1992 Olympic Games.

There are two main differences between the BExPlot and the Bagplot. The first one is that the reference point (centre) for the BExPlot is the sample mean, while for the Bagplot it is a bivariate median. The second is that if an observation is flagged as a possible outlier in the BExPlot, there is some univariate projection of the data for which that observation would be also flagged as a possible outlier in an eboxplot. This is because the expectile depth satisfies the strong projection property, see (12) and the comment below, together with the fact that in the eboxplot and BExPlot possible outliers are found dilating either the box (eboxplot) or the bag (BExPlot) by a factor of 4 from the centre. Conversely, and unlike in the Bagplot, all points marked as possible outliers in the eboxplot of one of the marginals, will be marked as possible outliers in the BExPlot.

Fig. 4 left represents the BExPlot of the time in the 200m race (horizontal axis) and the distance of the long jump (vertical axis) for the 26 female athletes that took part in the Heptathlon of the 1992 Barcelona Olympic Games. At margins we represent the eboxplots of the times at the 200m race and the long jump. The shaded region in the BExPlot is the 0.15-expectile region, whose horizontal and vertical projections are the boxes of the eboxplots presented at margins. Only one point is out of the barrier and appears marked as possible outlier. It corresponds to an athlete marked as possible outlier in the long jump eboxplot.

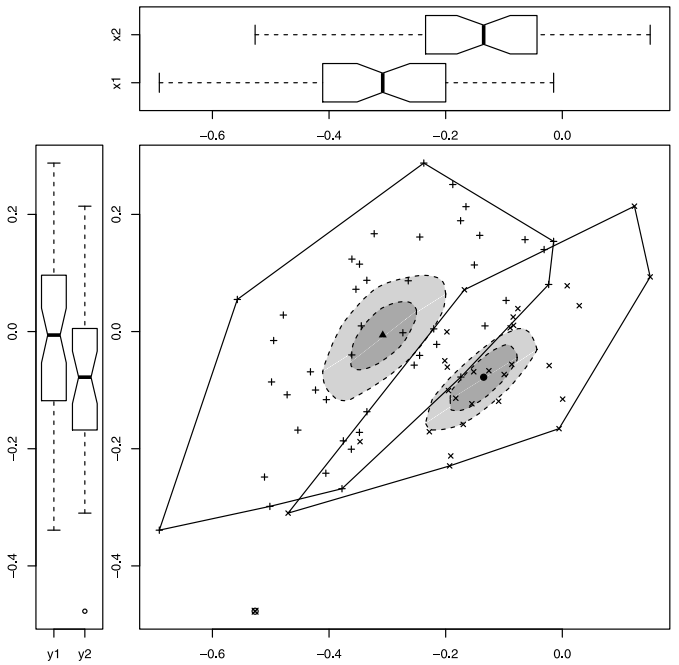


Fig. 5. BExPlots with confidence regions for the AFH antigen versus activity in Hemophilia A carriers (+) and non-carriers (×).

The Bagplot of the Barcelona92 dataset plotted with [48], is represented in Fig. 4 right. In this chart, two points are marked as possible outliers and the plots at margins are the qboxplots of the 200m race and the long jump.

The source code for the BExPlot in R is available on the GitHub repository https://github.com/icascos/expdepth. Two functions for it can be found there, `BExPlot` plots a fast approximation of the BExPlot, while `exactBExPlot` plots it with exact expectile regions computed with the algorithm described in Appendix A.1.

5.2. BExPlots with confidence regions

In order to provide a better impression of the data distribution, the contour of an expectile region that serves as an approximate confidence region on the mean, and whose level has been obtained under the assumption of normality, can be incorporated to the BExPlot. Together with it, approximate confidence intervals on the mean can be represented in the eboxplots at margins as notches in the boxes. Similarly, a confidence region on the halfspace depth median (deepest point with respect to the halfspace depth) was incorporated to the Bagplot in [40].

The approximate confidence region on the mean that we consider is the empirical expectile region of a level $\alpha(\beta, n, d)$ that depends on the confidence level β, sample size n, and dimension ($d = 2$ for the BExPlot and $d = 1$ for the eboxplot), but not on the data. Specifically, we compute the level of the region whose coverage probability for the sample mean of a normal sample of size n is the desired confidence level, that is, $\Pr\left((\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)/n \in \mathrm{ED}^{\alpha(\beta,n,d)}(\boldsymbol{X})\right) = \beta$, where $\boldsymbol{X}$ is any normal d-variate random vector and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ a random sample of size n taken from it.

In order to obtain the desired level, observe that the radius of the ball centred at the origin and containing the sample mean of n independent d-variate standard normal random vectors with probability β is $r = \left(F_d^{-1}(\beta)/n\right)^{1/2}$, where F_d^{-1} is the quantile function of a chi-squared random variable with d degrees of freedom. Finally, it matches the expectile region whose level α is the solution to the equation built from (19) substituting $e_\alpha(X)$ by $-r$,

$$\frac{\phi(-r)}{-r} + \Phi(-r) = \frac{\alpha}{2\alpha - 1}.$$

Notice that since the level of the expectile regions ranges between 0 and 1/2, we must consider the negative of the radius. For $d = 1$, we compute this way the level α that determines the notches in the eboxplot, while for $d = 2$, the level α that determines the approximate (bivariate) confidence region presented together with the BExPlot is computed.

In Fig. 5 we present a BExPlot for the *Hemophilia* dataset from Pokotylo et al. [37]. 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 × in the chart).

The notches in the eboxplots at margins represent approximate 95% confidence intervals on the respective means, while each one of the dark grey regions in the two BExPlots is the respective approximate 95% confidence region on

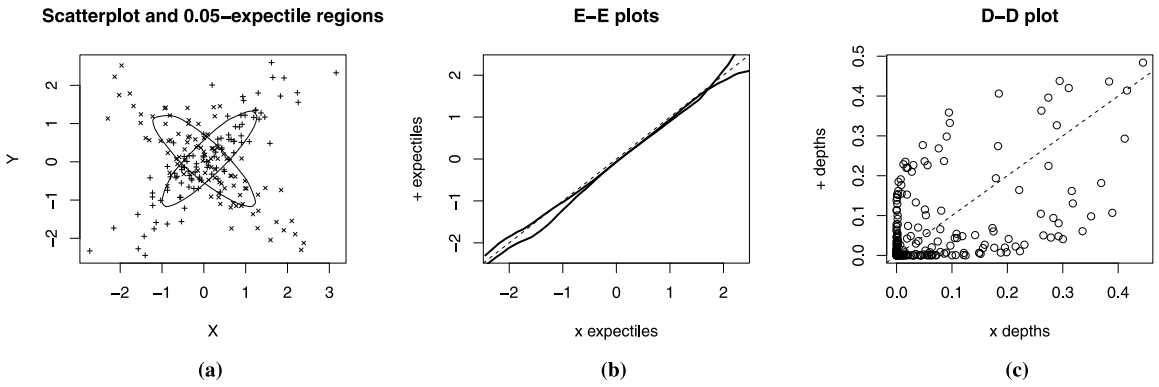


Fig. 6. (a) Scatterplot of two simulated datasets with their respective expectile regions of level 0.05; (b) E–E plots of first and second components of the two datasets; (c) expectile depth D–D plot of the two datasets.

the bivariate mean computed as described above. The dot marked as a possible outlier at both, the BExPlot and the eboxplot of variable AHF antigen (y), corresponds to a non-carrier of Hemophilia A. The two confidence regions are well separated. In fact, both samples can be assumed to have been drawn from normal populations (the respective p-values at the multivariate normality test in [47] are 0.7565 and 0.9586) whose means are not equal, since the p-value of the Hotelling's two-sample T-squared test on the equality of means is approximately 1.562×10^{-12}.

5.3. E–E plots and D–D plots

Schnabel and Eilers [42] proposed E–E (expectile–expectile) plots as a graphical technique to compare a univariate sample with a theoretical distribution. They claim that E–E plots constitute a less noisy alternative to the classical Q–Q plots, which is due to the continuity of the expectiles on the parameter α, see Property (ix) in Section 2.1. A multivariate depth-based extension of the Q–Q plots was proposed by Liu et al. [30] under the name of D–D (depth–depth) plots. These D–D plots serve to compare two multivariate samples by plotting the depth of each data point of the pooled dataset with respect to one sample versus the depth with respect to the other sample. When both samples have been drawn from the same distribution, all the points of the D–D plot lie close to the bisector of the first quadrant. Next we use E–E plots to compare two univariate samples and D–D plots built for the expectile depth.

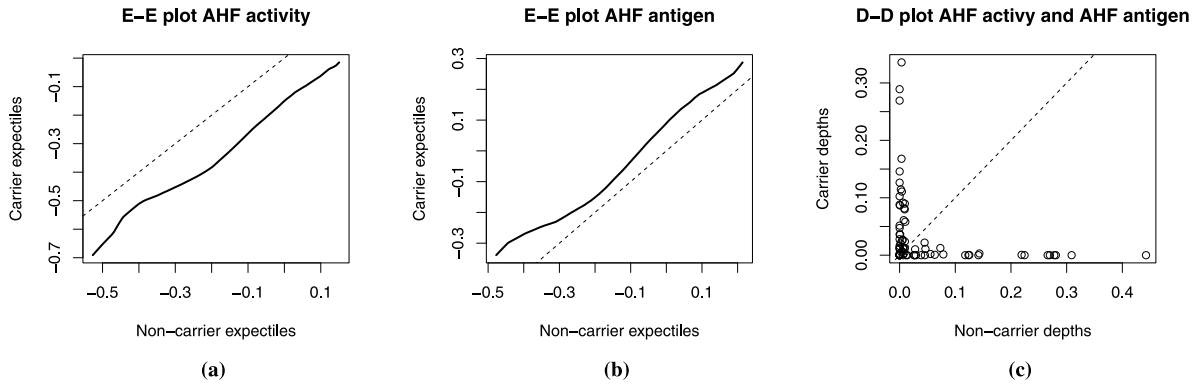
We have simulated two samples of 100 observations each from bivariate normal distributions with standard normal marginals, and respective correlations equal to 0.85 and -0.85. The scatterplot of the simulated data is presented in Fig. 6(a), where the sample of positively correlated data is marked as $+$ and that of negatively correlated data as $\times$. The contour of the expectile region of level 0.05 of each dataset is also presented. In Fig. 6(b) the E–E plots of the first (and second) components of the two samples are presented. Since both first components (and both second components) follow the same distribution (standard normal), their E–E plots adjust well to the straight line through the origin with slope 1. Finally, Fig. 6(c) represents the D–D plot (for the expectile depth, obtained using the algorithm in Appendix A.2) with the 200 observations. Despite there is some observation which is central with regard to both clouds, showing a high value for both depths, the D–D plot does not adjust to the straight line through the origin with slope 1, showing that the samples do not share the same distribution.

In Fig. 7 we have represented the E–E plots of the AHF activity (a) and AHF antigen (b) of carriers versus non-carriers in the *Hemophilia* dataset. The E–E plot in Fig. 7(a) lies below the straight line through the origin with slope 1, because the non-carriers have greater AHF activity values than the carriers. As for AHF antigen, see Fig. 7(b), carriers exhibit greater values than non-carriers, and thus the E–E plot lies above the reference line. Finally, the D–D plot in Fig. 7(c) shows that the joint distribution of AHF activity and antigen in carriers is different from the one in non-carriers. Observe that there are many points close to both of the axes, representing observations with low depth with respect to one of the samples, despite some of them are somehow central with respect to the other sample.

6. Highlights and conclusions

The main achievement in this paper is the comprehensive 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 the univariate expectile function and a summary of its main properties. Some emphasis is placed on the inverse expectile function and the empirical expectiles and inverse expectiles.

Expectile regions are defined as intersections of halfspaces determined by univariate expectiles. Due to the fact that expectiles are positively homogeneous and subadditive functions for specific values of their level α, we present expectile regions as compact convex sets whose support functions are given in terms of expectiles of univariate projections of the



(a) (b) (c)

Fig. 7. (a) E–E plot of AHF activity (carriers vs. non-carriers); (b) E–E plot of AHF antigen (carriers vs. non-carriers); (c) expectile depth D–D plot of the *Hemophilia* dataset (carriers vs. non-carriers).

data. Some basic properties of the expectile regions are discussed, and it is shown that the family of expectile regions characterizes a distribution. Together with the population expectile regions, we present their empirical counterparts, 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 relevance of some of those properties to the topology of the expectile regions. We present next the sample expectile depth function and show its uniform consistency. A second algorithm, this one to compute the expectile depth of a point with respect to a multivariate dataset, is presented.

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. The BExPlot is centred at the mean of a dataset and we represent an approximate confidence region on the mean together with it based on normality assumptions. We understand that it can be useful to illustrate the Hotelling's T-squared test on the comparison of two multivariate means or classical MANOVA procedures to compare means under normality assumptions, with the obvious caution that it represents an approximate confidence region on each individual mean.

CRediT authorship contribution statement

Ignacio Cascos: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation Data curation, Writing, Visualization, Supervision. **Maicol Ochoa:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing, Visualization.

Acknowledgements

This manuscript has benefited from the insightful and constructive comments of three referees and the very professional efforts of the JMVA editors, who actively took part in its improvement. We would like to thank Prof. Paul Eilers for presenting his work on expectiles to us at a seminar in the Carlos III University. This research was partially supported by the Spanish Ministry of Science and Innovation under grant ECO2015-66593-P.

Appendix A. Algorithms

A.1. Algorithm for computing the extreme points of the bivariate expectile regions

(R source code available on the GitHub repository https://github.com/icascos/expdepth as function `exactexp`)

- *Input:* Data points $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$, $i = 1, \ldots, n$ and depth level $0 < \alpha \leq 1/2$.
- *Output:* Extreme points of ED_n^α supported by half-planes with outer normal $(\cos\beta, \sin\beta)$ with $0 \leq \beta \leq \pi$ (northern boundary).

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 rth row of **ANG** contains the rth 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 (i, index of the first point that defines the angle), and all rows with the same first and second entries are ordered in terms of their third entries (j, index of the second point that defines the angle).

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. The first row in **ANG** has thus entries $\mathbf{ANG}_{0,1} = 0$ (the angle), while entries $\mathbf{ANG}_{0,2}$ and $\mathbf{ANG}_{0,3}$ (the point indices) 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 (14). 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 (14) 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 interchanges 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 (14).

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 sth 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 an 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. In the array **R**, interchange the values at positions $\mathbf{ANG}_{k,2}$ and $\mathbf{ANG}_{k,3}$.

Step 12. While $k < \binom{n}{2}$, set $k \leftarrow k + 1$ and go to Step 7.

Remember that current algorithm 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.

In order to discuss the complexity of the algorithm notice that:

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

(ii) The first computation of a univariate expectile, performed in Step 4., can be done using (8) as an alternative to the repeated weighted averaging numerical 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 for the empirical cdf evaluated on the expectile were to be considered, the complexity would increase at most up to n^2.

(iii) 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 computing the bivariate expectile depth

(R source code available on the GitHub repository https://github.com/icascos/expdepth as function `expdepth`)

- *Input:* Data points $\boldsymbol{x}_i \in \mathbb{R}^2$, $i = 1, \ldots, n$ and point $\boldsymbol{y} \in \mathbb{R}^2$ whose expectile depth is to be computed.
- *Output:* Depth $\mathrm{ED}_n(\boldsymbol{y})$ of $\boldsymbol{y}$ with respect to the dataset $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.

Step 1. Centre all data points with respect to point $\boldsymbol{y}$.

Step 2. Compute the sum of all observations and store it in bivariate vector **sdata** whose first component is the sum of all first components and the second component is the sum of the second components of the data points.

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

Step 4. Reflect through the origin those points whose angles are between $\frac{\pi}{2}$ and $\frac{3\pi}{2}$ by subtracting π 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. Compute bivariate vectors **spos** and **sneg** as the sums of data points that respectively lie in the halfplane with inner and outer normal $(\cos(\gamma_1 + \pi/2), \sin(\gamma_1 + \pi/2))$ and set $i = 1$.
Step 7. Set $\boldsymbol{u} = (\cos(\gamma_i + \pi/2), \sin(\gamma_i + \pi/2))$
Step 8. Compute the minimum between $\langle \mathbf{sdata}, \boldsymbol{u}\rangle / \langle \mathbf{spos}, \boldsymbol{u}\rangle$ and $\langle \mathbf{sdata}, -\boldsymbol{u}\rangle / \langle \mathbf{sneg}, -\boldsymbol{u}\rangle$ store this value in an array called **MIN**.
Step 9. While $i < n$, set $i \to i + 1$ and update $\boldsymbol{u} = (\cos(\gamma_i + \pi/2), \sin(\gamma_i + \pi/2))$ together with **spos** and **sneg**. Observe that the latter vectors are updated by adding or subtracting the point whose coordinates are $(\mathbf{ANG}_{i,1}, \mathbf{ANG}_{i,2})$, according to the tag in the fourth column of matrix **ANG** and go to Step 7.
Step 10. With the minimum of those values stored in **MIN** compute $(2 - \min \mathbf{MIN})^{-1}$ as in (17) and return that value.

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 **spos** and **sneg** involves at most the addition and subtraction of two points, so its complexity is $O(n)$ and does not affect the overall complexity of the current algorithm, which remains $O(n \log n)$.

Appendix B. Mathematical proofs

B.1. Derivation of Eq. (3)

The expectile can be extracted form (2) in order to obtain an expression for it that depends on the mean of the random variable and the expectation of the positive part of $X - e_\alpha(X)$,

$$\begin{aligned} e_\alpha(X) &= \mathrm{E}X + \frac{2\alpha - 1}{1-\alpha}\mathrm{E}(X - e_\alpha(X))_+ = \mathrm{E}X + \frac{2\alpha - 1}{1-\alpha}\int_{F_X(e_\alpha(X))}^1 \left(F_X^{-1}(t) - e_\alpha(X)\right)\,\mathrm{d}t \\ &= \mathrm{E}X + \frac{2\alpha - 1}{1-\alpha}\left[\int_{F_X(e_\alpha(X))}^1 F_X^{-1}(t)\,\mathrm{d}t - e_\alpha(X)\Big(1 - F_X\big(e_\alpha(X)\big)\Big)\right]. \end{aligned}$$

Observe that since the expectile of X appears in both of the left and the right hand side of the expression above, it can be cleared, leaving the cdf of X evaluated at it on the right-hand side,

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

B.2. Proof of the general properties of the expectile regions in Section 3.1

(i) States that the most central expectile region is the singleton formed by the expectation. This follows from the linearity of the expectation and the fact that the expectile of level 1/2 is the mean (Property (i) in Section 2.1). In fact, for any random vector $\boldsymbol{X}$ with finite first moment and $\boldsymbol{u} \in \mathbb{R}^d$, $e_{1/2}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) = \mathrm{E}\langle \boldsymbol{X}, \boldsymbol{u}\rangle = \langle \mathrm{E}\boldsymbol{X}, \boldsymbol{u}\rangle$, and the only intersection point of all halfspaces $\{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{u}\rangle \le \langle \mathrm{E}\boldsymbol{X}, \boldsymbol{u}\rangle\}$ is $\mathrm{E}\boldsymbol{X}$.

(ii) The nesting property follows from the parameter monotonicity of the expectiles (Property (x) in Section 2.1). Take any $0 < \alpha \le 1/2$, if $\boldsymbol{x} \in \mathrm{ED}^\alpha(\boldsymbol{X})$, then for every $\boldsymbol{u} \in \mathbb{R}^d$, $\langle \boldsymbol{x}, \boldsymbol{u}\rangle \le e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$. Take now $0 < \beta < \alpha$, then $1-\beta > 1-\alpha$, and $e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \le e_{1-\beta}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle)$. Since the inequalities hold for every $\boldsymbol{u} \in \mathbb{R}^d$, we have $\boldsymbol{x} \in \mathrm{ED}^\alpha(\boldsymbol{Y})$.

(iii),(iv) The convexity and closedness of the expectile regions follow from their construction as intersection of closed halfspaces in (10). In order so show their compacity, it is enough to check that $\mathrm{ED}^\alpha(\boldsymbol{X})$ is bounded, which holds true since it is contained in the d-dimensional parallelotope $\mathrm{ED}^\alpha(\boldsymbol{X}) \subseteq \{\boldsymbol{x} \in \mathbb{R}^d : e_\alpha(X_i) \le x_i \le e_{1-\alpha}(X_i)\}$, where $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{X} = (X_1, \ldots, X_d)$.

(v) The affine equivariance of the expectile regions follows from the translation equivariance of the expectiles (Property (iii) in Section 2.1) and their characterization in terms of ther support function, see (12). Consider any $\boldsymbol{A} \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$, $h(\mathrm{ED}^\alpha(\boldsymbol{AX}+\boldsymbol{b}), \boldsymbol{u}) = e_{1-\alpha}(\langle \boldsymbol{AX}+\boldsymbol{b}, \boldsymbol{u}\rangle) = e_{1-\alpha}(\langle \boldsymbol{AX}, \boldsymbol{u}\rangle) + \langle \boldsymbol{b}, \boldsymbol{u}\rangle = e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{A}^\top\boldsymbol{u}\rangle) + \langle \boldsymbol{b}, \boldsymbol{u}\rangle = h(\mathrm{ED}^\alpha(\boldsymbol{X}), \boldsymbol{A}^\top\boldsymbol{u}) + \langle \boldsymbol{b}, \boldsymbol{u}\rangle = h(\boldsymbol{A}\mathrm{ED}^\alpha(\boldsymbol{X}), \boldsymbol{u}) + \langle \boldsymbol{b}, \boldsymbol{u}\rangle = h(\boldsymbol{A}\mathrm{ED}^\alpha(\boldsymbol{X}) + \boldsymbol{b}, \boldsymbol{u})$.

(vi) The monotonicity of the expectile regions follows from the monotonicity of the univariate expectiles (Property (v) in Section 2.1). If $\boldsymbol{X} \le \boldsymbol{Y}$ componentwisely, then for every $\boldsymbol{u} \in \mathbb{R}^d_+$ it holds $\langle \boldsymbol{X}, \boldsymbol{u}\rangle \le \langle \boldsymbol{Y}, \boldsymbol{u}\rangle$, or equivalently, $\langle \boldsymbol{X}, \boldsymbol{u}\rangle \ge \langle \boldsymbol{Y}, \boldsymbol{u}\rangle$ for every $\boldsymbol{u} \in \mathbb{R}^d_-$. Hence, for $0 < \alpha \le 1/2$ and $\boldsymbol{u} \in \mathbb{R}^d_-$ we have $h(\mathrm{ED}^\alpha(\boldsymbol{X}), \boldsymbol{u}) = e_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) \ge e_{1-\alpha}(\langle \boldsymbol{Y}, \boldsymbol{u}\rangle) = h(\mathrm{ED}^\alpha(\boldsymbol{Y}), \boldsymbol{u})$, or equivalently $\mathrm{ED}^\alpha(\boldsymbol{Y}) \oplus \mathbb{R}^d_+ \subseteq \mathrm{ED}^\alpha(\boldsymbol{X}) \oplus \mathbb{R}^d_+$.

(vii) The subadditivity of the expectile regions follows from the subadditivity of the univariate expectiles (Property (vii) in Section 2.1). If $\boldsymbol{z} \in \mathrm{ED}^\alpha(\boldsymbol{X} + \boldsymbol{Y})$ for some $0 < \alpha \le 1/2$, then for all $\boldsymbol{u} \in \mathbb{S}^{d-1}$ it holds $\langle \boldsymbol{z}, \boldsymbol{u}\rangle \le e_{1-\alpha}\langle \boldsymbol{X} + \boldsymbol{Y}, \boldsymbol{u}\rangle \le e_{1-\alpha}\langle \boldsymbol{X}, \boldsymbol{u}\rangle + e_{1-\alpha}\langle \boldsymbol{Y}, \boldsymbol{u}\rangle = h(\mathrm{ED}^\alpha(\boldsymbol{X}), \boldsymbol{u}) + h(\mathrm{ED}^\alpha(\boldsymbol{Y}), \boldsymbol{u}) = h(\mathrm{ED}^\alpha(\boldsymbol{X}) + \mathrm{ED}^\alpha(\boldsymbol{Y}), \boldsymbol{u})$, so $\boldsymbol{z} \in \mathrm{ED}^\alpha(\boldsymbol{X}) + \mathrm{ED}^\alpha(\boldsymbol{Y})$. □

B.3. Proof of Proposition 1

Fix $\boldsymbol{u} \in \mathbb{S}^{d-1}$ and consider a sequence α_n such that $\lim_n \alpha_n = \alpha$ for some $0 < \alpha \leq 1/2$, since the expectiles are continuous with respect to the parameter α, see Property (ix) 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 [43, Th. 1.8.12], the pointwise and uniform convergence of support functions on $\mathbb{S}^{d-1}$ are equivalent, thus $\lim_n d_H(\mathrm{ED}^{\alpha_n}(\boldsymbol{X}), \mathrm{ED}^{\alpha}(\boldsymbol{X})) = 0$. □

B.4. Proof of Proposition 2

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

Since each $\boldsymbol{y}_n$ does not lie in $\mathrm{ED}^{\beta}(\boldsymbol{X})$ and by the strict parameter monotonicity of expectiles, see Property (x) 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 $\boldsymbol{y} \in \mathrm{ED}^{\alpha}(\boldsymbol{X})$. □

B.5. Proof of Proposition 3

Following Holzmann and Klar [23, Th. 2], for any fixed $\boldsymbol{u} \in \mathbb{S}^{d-1}$ and $\alpha \in (0, 1/2)$, the (sequence on n of) 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, $\{h(\mathrm{ED}_n^{\alpha}, \boldsymbol{u}) : n \geq 1\}$ converge a.s. to $h(\mathrm{ED}^{\alpha}(\boldsymbol{X}), \boldsymbol{u})$.

According to Molchanov [32, 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, $\{\mathrm{ED}_n^{\alpha} : n \geq 1\}$ converges a.s. to $\mathrm{ED}^{\alpha}(\boldsymbol{X})$ in the Hausdorff metric.

Following Dyckerhoff [14, Th. 4.7], the strict monotonicity condition of the expectile depth shown in part (b) of Proposition 7 and the almost sure convergence of $\{\mathrm{ED}_n^{\alpha} : n \geq 1\}$ in the Hausdorff metric for any fixed α imply the almost sure uniform convergence of $\{\mathrm{ED}_n^{\alpha} : n \geq 1\}$ in the Hausdorff metric on a compact set $\alpha \in I \subseteq (0, 1/2)$. In conclusion,

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

B.6. Proof of Proposition 5

An explicit expression for the expectile depth can be obtained from its definition as a supremum, the formula for the inverse expectile function in (4), and elementary algebra,

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

B.7. Proof of Proposition 6

After Property (ii) in Section 4.1, the map $\boldsymbol{y} \mapsto \mathrm{ED}(\boldsymbol{y}; \boldsymbol{X})$ is continuous. Together with Proposition 3, Dyckerhoff [14, Th. 4.6] has shown that this implies the almost sure uniform convergence of the sequence of depths $\{\mathrm{ED}_n(\,\cdot\,; \boldsymbol{X}) : n \geq 1\}$. □

B.8. Proof of Proposition 7

(i) Without loss of generality, consider a random vector centred at the origin, $\mathrm{E}\boldsymbol{X} = \mathbf{0}$, 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 the 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 for $\mathrm{ED}(\lambda\boldsymbol{x}; \boldsymbol{X})$.

(ii) As shown in [14, Th. 3.2], part (ii) is equivalent to the continuity of the map $\alpha \mapsto \mathrm{ED}^{\alpha}(\boldsymbol{X})$ with respect to the Hausdorff metric proved in Proposition 1. □

References

[1] F. Bellini, B. Klar, A. Müller, Expectiles, omega ratios and stochastic ordering, Methodol. Comput. Appl. Probab. 20 (2018) 855–873.
[2] F. Bellini, B. Klar, A. Müller, E. Rosazza Gianin, Generalized quantiles as risk measures, Insurance Math. Econom. 54 (2014) 41–48.
[3] J. Breckling, R. Chambers, M-quantiles, Biometrika 75 (1988) 761–771.
[4] J. Breckling, P. Kokic, O. Lübke, A note on multivariate M-quantiles, Statist. Probab. Lett. 55 (2001) 39–44.
[5] I. Cascos, The expected convex hull trimmed regions of a sample, Comput. Statist. 22 (2007) 557–569.
[6] I. Cascos, Data depth: multivariate statistics and geometry, in: W. Kendall, I. Molchanov (Eds.), New Perspectives in Stochastic Geometry, Oxford University Press, 2010, pp. 398–423.
[7] I. Cascos, M. López-Díaz, Control charts based on parameter depths, Appl. Math. Model. 53 (2018) 487–509.
[8] I. Cascos, I. Molchanov, Multivariate risks and depth-trimmed regions, Financ. Stoch. 8 (2007) 531–552.
[9] G. Choquet, Theory of capacities, Ann. Inst. Fourier 5 (1954) 131–295.
[10] H. Cramér, H. Wold, Some theorems on distribution functions, J. London Math. Soc. s1- 11 (1936) 290–294.
[11] A. Daouia, D. Paindaveine, From halfspace M-depth to multiple output expectile regression, 2019, arXiv e-prints, arXiv:1905.12718.
[12] R. Dyckerhoff, Computing zonoid trimmed regions of bivariate data sets, in: J. Bethlehem, P. Heijden (Eds.), Proceedings in Computational Statistics, COMPSTAT 2000, Physica–Verlag, Heidelberg, 2000, pp. 295–300.
[13] R. Dyckerhoff, Data depths satisfying the projection property, All. Stat. Arch. 88 (2004) 163–190.
[14] R. Dyckerhoff, Convergence of depths and depth-trimmed regions, 2017, arXiv e-prints, arXiv:1611.08721v2.
[15] R. Dyckerhoff, K. Mosler, Weighted-mean trimming of multivariate data, J. Multivariate Anal. 102 (2011) 405–421.
[16] R. Dyckerhoff, K. Mosler, Weighted-mean regions of a probability distribution, Statist. Probab. Lett. 82 (2012) 318–325.
[17] A. Eberl, B. Klar, Expectile based measures of skewness, 2019, arXiv e-prints, arXiv:1908.08243.
[18] P. Eilers, Expectile contours and data depth, in: A.W. Bowman (Ed.), Proceedings of the 25th International Workshop on Statistical Modelling, 2010, pp. 167–172.
[19] A. Gelman, X.-L. Meng, A note on bivariate distributions that are conditionally normal, Amer. Stat. 45 (1991) 125–126.
[20] E. Giorgi, A.J. McNeil, On the computation of multivariate scenario sets for the skew-t and generalized hyperbolic families, Comput. Statist. Data Anal. 100 (2016) 205–220.
[21] T. Gneiting, Making and evaluating point forecasts, J. Amer. Statist. Assoc. 106 (2011) 746–762.
[22] K. Herrmann, M. Hofert, M. Mailhot, Multivariate geometric expectiles, Scand. Actuar. J. 7 (2018) 629–659.
[23] H. Holzmann, B. Klar, Expectile asymptotics, Electron. J. Stat. 10 (2016) 2355–2371.
[24] M.C. Jones, Expectiles and M-quantiles are quantiles, Statist. Probab. Lett. 20 (1994) 149–153.
[25] G. Koshevoy, K. Mosler, Zonoid trimming for multivariate distributions, Ann. Statist. 25 (1997) 1998–2017.
[26] G. Koshevoy, K. Mosler, Lift zonoids, random convex hulls and the variability of random vectors, Bernoulli 4 (1998) 377–399.
[27] V. Krätschmer, H. Zähle, Statistical inference for expectile-based risk measures, Scand. J. Stat. 44 (2017) 425–454.
[28] R.Y. Liu, On a notion of data depth based on random simplices, Ann. Statist. 18 (1990) 405–414.
[29] R.Y. Liu, Control charts for multivariate processes, J. Amer. Statist. Assoc. 90 (1995) 1380–1387.
[30] R.Y. Liu, J.M. Parelius, K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh), Ann. Stat. 27 (1999) 783–858.
[31] V. Maume-Deschamps, D. Rullière, K. Said, Multivariate extensions of expectile risk measures, Depend. Model. 5 (2016) 20–44.
[32] I. Molchanov, Theory of Random Sets, second ed., Springer, London, 2017.
[33] K. Mosler, Multivariate Dispersion, Central Regions, and Depth, in: The lift zonoid approach, Lecture Notes in Statistics, vol. 165, Springer, Berlin, 2002.
[34] A. Müller, Orderings of risks: A comparative study via stop-loss transforms, Insurance Math. Econom. 17 (1996) 215–222.
[35] A. Müller, D. Stoyan, Comparison Methods for Stochastic Models and Risks, Wiley, 2002.
[36] W. Newey, J. Powell, Asymmetric least squares estimation and testing, Econometrika 55 (1987) 819–847.
[37] O. Pokotylo, P. Mozharovskyi, R. Dyckerhoff, S. Nagy, ddalpha: Depth-based classification and calculation of data depth 1.3.9, 2019, https://cran.r-project.org/web/packages/ddalpha/index.html.
[38] P.J. Rousseeuw, I. Ruts, Algorithm AS 307: Bivariate location depth, J. Roy. Statist. Soc. C 45 (1996) 516–526.
[39] P.J. Rousseeuw, I. Ruts, The depth function of a population distribution, Metrika 49 (1999) 213–244.
[40] P.J. Rousseeuw, I. Ruts, J.W. Tukey, The bagplot: A bivariate boxplot, Amer. Stat. 53 (1999) 382–387.
[41] I. Ruts, P.J. Rousseeuw, Computing depth contours of bivariate point clouds, Comput. Statist. Data Anal. 23 (1996) 153–168.
[42] S. Schnabel, P. Eilers, Optimal expectile smoothing, Comput. Statist. Data Anal. 53 (2009) 4168–4177.
[43] R. Schneider, Convex Bodies: The Brunn–Minkowski Theory, in: Encyclopedia of Mathematics and Its Applications, vol. 44, Cambridge University Press, Cambrigde, 1993.
[44] M. Shaked, G. Shantikumar, Stochastic Orders, in: Springer Series in Statistics, 2007.
[45] F. Sobotka, S. Schnabel, L.S. Waltrup, P. Eilers, G. Kauermann, expectreg: Expectile and quantile regression, 2014, https://github.com/cran/expectreg.git, Github read-only mirror of the CRAN R package repository.
[46] J.W. Tukey, Mathematics and the picturing of data, in: R.D. James (Ed.), Proceedings of the International Congress of Mathematitians (Vancouver, B.C. 1974), in: Canad. Math. Congress, vol. 2, Que, Montreal, 1975, pp. 523–531.
[47] J.A. Villasenor-Alva, E. González-Estrada, A generalization of Shapiro–Wilk's test for multivariate normality, Comm. Statist. Theory Methods 38 (2009) 1870–1883.
[48] H.P. Wolf, aplpack: Another plot package 1.3.2, 2018, https://cran.r-project.org/web/packages/aplpack/index.html.
[49] X. Xu, J. Sun, S. Nie, H. Li, Y. Kong, M. Liang, J. Hou, X. Huang, D. Li, T. Ma, J. Peng, et al., Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China, Nat. Med. 26 (2020) 1193–1195.
[50] J. Ziegel, Coherence and elicitability, Math. Finance 26 (2016) 901–918.
[51] Y. Zuo, R. Serfling, General notions of statistical depth function, Ann. Statist. 28 (2000) 461–482.