Working paper

2022-07

Statistics and Econometrics
ISSN 2387-0303

Data depth and multiple output regression, the distorted M-quantiles approach

Maicol Ochoa, Ignacio Cascos

uc3m | Universidad Carlos III de Madrid

Data depth and multiple output regression, the distorted M-quantiles approach

Maicol Ochoa and Ignacio Cascos
Department of Statistics, Universidad Carlos III de Madrid

July 14, 2022

Abstract

For a univariate distribution, its M-quantiles are obtained as solutions to asymmetric minimization problems dealing with the distance of a random variable to a fixed point. The asymmetry refers to the different weights for the values of the random variable at either side of the fixed point. We focus on M-quantiles whose associated losses are given in terms of a power. In this setting, the classical quantiles are obtained for the first power, while the expectiles correspond to quadratic losses. The M-quantiles considered here are computed over distorted distributions, which allows to tune the weight awarded to the more central or peripheral parts of the distribution. These distorted M-quantiles are used in the multivariate setting to introduce novel families of central regions and their associated depth functions, which are further extended to the multiple output regression setting in the form of conditional regression regions and conditional depths.

Keywords: Bivariate depth algorithm; Data depth; Distortion function; Conditional regression region; M-quantiles.

1 Introduction

In multivariate statistics, the degree of centrality of a point $\boldsymbol{x}$ with respect to a multivariate probability distribution or data sample is known as the *depth* of $\boldsymbol{x}$. Since the seminal introduction of the halfspace regions (sets of points whose halfspace depth is, at least, some given value) by Tukey (1975) in the context of graphical representations of bivariate datasets, a large number of notions of depth have been proposed and studied in the statistical literature. We highlight here the simplicial depth introduced by Liu (1990) proving some properties that have become standard requirements for subsequent depth proposals, the first formal study of the halfspace depth by Rousseeuw and Ruts (1999), and the introduction of the zonoid depth by Koshevoy and Mosler (1997). See also Cascos (2010), Liu *et al.* (1999), Zuo and Serfling (2000a,b) for some systematic reviews of depth functions, central or depth regions, and their properties.

There has been recent interest in the use of expectiles (and more generally M-quantiles) in the construction of central regions and depth functions, see Cascos and Ochoa (2021), Daouia and Paindaveine (2019). These expectiles, first introduced by Newey and Powell (1987) in the context of linear regression, are the solution to an asymmetrically weighted least squares minimization problem. Likewise, the M-quantiles constitute a natural generalization of the expectiles and the standard quantiles that keep the asymmetric weights, but whose loss function is not necessarily quadratic or an absolute value, see Breckling and Chambers (1988).

As it will be shown, the M-quantiles of a distribution can be tuned, in particular robustified, by introducing a distortion function which adjusts the weights awarded to the values in the support of the distribution in terms of the quantile they correspond to. The so-called distorted M-quantiles turn out to be the solution to an equation involving Choquet expectations with respect to a (distorted) non-additive probability. Throughout this manuscript, specific distortions that downweight (or upweight) the outermost observations from a dataset are considered. Based on them, distorted M-quantile regions and their associated depth functions, generalizing those in Cascos and Ochoa (2021), Daouia and Paindaveine (2019), are obtained.

In the last part of the manuscript, we follow Daouia and Paindaveine (2019), Hallin *et al.* (2010, 2015), Merlo *et al.* (2022), to present the notion of distorted M-quantile conditional regression region in the context of multiple output linear regression. The single output (distorted) M-quantile regression resembles the classical quantile regression introduced by Koenker and Basset (1978), see also Koenker (2005), while its extension to multi-output models is based on univariate projections of the response variable. In this setting, we also consider a notion of conditional data depth for the output of a regression model given the regressors.

The M-quantile conditional regression regions are intended to serve as a descriptive tool when analyzing the dependent variables given specific values of the predictors. Through some examples with both real and simulated datasets, we illustrate the robustness of the distorted M-quantile regression models against the presence of extreme points, as well as that of the corresponding regions and depth when using appropriate distortion functions.

This manuscript is organized as follows: Section 2 contains a summary of the preliminary concepts, where the notions of non-additive probabilities and Choquet expectations are used to introduce the concept of M-quantiles with power loss function. In Section 3 some relevant properties of the distorted M-quantiles, including an inversion formula and brief discussion of the sample distorted expectiles, are presented. Section 4 is devoted to the notion of distorted M-quantile depth, which is obtained in terms of the distorted M-quantile regions. In Section 5, we use a projection-based M-quantile regression technique for multi-output models from which we obtain the M-quantile conditional regression regions. Finally, a list of concluding remarks and future work is displayed in Section 6. Appendix A is placed at the end of the manuscript describing an algorithm to compute the bivariate M-quantile depth.

2 Preliminaries

2.1 M-quantiles with power loss function

Consider a general nonatomic probability space $(\Omega, \mathcal{F}, P)$ and a random variable X with finite moment of order $r \geq 1$ defined on it. For any $0 < \alpha < 1$, the *power M-quantile* of X of order r and level α is the solution to the minimization problem

$$q_\alpha^{(r)}(X) = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}[\alpha(X-\theta)_+^r + (1-\alpha)(X-\theta)_-^r], \tag{1}$$

where $x_+ = \max\{x, 0\}$ and $x_- = (-x)_+$ for any $x \in \mathbb{R}$.

The M-quantiles with power loss function defined in (1) satisfy appealing properties like translation equivariance, positive homogeneity, and being uniquely defined for $r > 1$, see Bellini *et al.* (2014) for the proofs and further details.

Notice that when $r = 1$, we obtain the classical quantiles for which the uniqueness is only achieved if X has a continuous distribution with strictly increasing cdf, while for $r = 2$, we obtain the *expectiles*, which we denote by $e_\alpha(X) = q_\alpha^{(2)}(X)$.

In order to relax the notation, hereafter we omit the order of the M-quantile (unless there is some need to emphasize it) and write $q_\alpha(X)$ for the M-quantile of X of order r, $q_\alpha^{(r)}(X)$.

Observe that the loss function in (1) is $\rho(x, \theta) = \alpha(x-\theta)_+^r + (1-\alpha)(x-\theta)_-^r$, and the asymmetric weights for the losses to the left and right of θ are respectively $1-\alpha$ and α. Differentiating ρ with respect to θ, we obtain the first order condition equation that renders the M-quantile of order $r > 1$ as the unique solution to the equation

$$\alpha \mathbb{E}(X-\theta)_+^{r-1} = (1-\alpha)\mathbb{E}(X-\theta)_-^{r-1}, \tag{2}$$

which only requires $\mathbb{E}|X|^{r-1}$ to be finite. Equation (2) can be also used for $r = 1$ with the agreement that $0^0 = 0$, and loosing the uniqueness in the solution, in order to compute the standard quantiles.

2.2 Non-additive probabilities

A *non-additive probability* over the reference measurable space $(\Omega, \mathcal{F})$ is any normalized and monotone function from the σ-algebra $\mathcal{F}$. That is, a function $T : \mathcal{F} \to \mathbb{R}$ such that $T(\emptyset) = 0$, $T(\Omega) = 1$, and $T(B) \leq T(C)$ if $B \subseteq C$, see Denneberg (1994), Molchanov (2017). The dual to T is the non-additive probability $\widetilde{T}(B) = 1 - T(B^c)$, where B^c is the complement of B.

Consider a measurable mapping X from $(\Omega, \mathcal{F})$ into $\mathbb{R}_+$ equipped with the Borel σ-algebra. The *asymmetric Choquet expectation* of X with respect to T is

$$\mathbb{CE}_T X = \int_0^\infty T(X \geq x)\,\mathrm{d}x, \tag{3}$$

where the term in the right appears in the form of a Riemann integral. It is known that the asymmmetric Choquet expectation with respect to a non-additive probability is a positive homogeneous ($\mathbb{CE}_T[\lambda X] = \lambda \mathbb{CE}_T X$ if $\lambda > 0$), monotone ($\mathbb{CE}_T X \leq \mathbb{CE}_T Y$ if $X \leq Y$), and translation equivariant ($\mathbb{CE}_T[b + X] = b + \mathbb{CE}_T X$ if $b \in \mathbb{R}$ with $X \geq \max\{0, -b\}$) operator, see Denneberg (1994, Prop. 5.1).

Any general $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$ can be written as $X = X_+ - X_-$ and its *symmetric Choquet expectation* is $\mathbb{CE}_T X = \mathbb{CE}_T X_+ - \mathbb{CE}_{\widetilde{T}} X_-$. Since we have restricted the asymmetric Choquet expectation to mappings into $\mathbb{R}_+$ and both Choquet expectations match on those nonnegative mappings, we can use the same notation for them. The symmetric Choquet expectation is positive homogeneous, monotone, and symmetric ($\mathbb{CE}_T[-X] = -\mathbb{CE}_T X$). For nonnegative mappings, $X \geq 0$, this symmetry results into the asymmetry-type relation $\mathbb{CE}_T[-X] = -\mathbb{CE}_{\widetilde{T}} X$.

2.3 Distortion functions

A *distortion function* is any non-decreasing function $g : [0, 1] \to [0, 1]$ such that $g(0) = 0$ and $g(1) = 1$. When a distortion function g acts over a probability measure P, we get what is known as a *distorted probability*, which is defined as the non-additive probability $P^*(B) = g(P(B))$ for $B \in \mathcal{F}$. Correspondingly, when a distortion function g acts on a cumulative distribution function (cdf) F, it transforms F into the new cdf $g(F)$. The dual distortion function of g is $\widetilde{g}(x) = 1 - g(1 - x)$ and it acts on P transforming it on the dual to P^*. Finally, we say that a distortion function is symmetric when $\widetilde{g} = g$, and in such a case the distorted probability also matches its dual. For further details, see Denneberg (1994, Chapter 2).

Particular instances of symmetric distortion functions are the identity function, whose associated (distorted) probability is P itself, the *trim distortion* of parameter $0 \le \beta < 1$

$$g_\beta(x) = \begin{cases} 0 & \text{if } 0 \le x \le \beta/2 \\ \frac{x-\beta/2}{1-\beta} & \text{if } \beta/2 < x < 1-\beta/2 \\ 1 & \text{if } 1-\beta/2 \le x \le 1 \end{cases},$$

and the *sigmoid distortion* of parameter $\delta > 0$ defined as

$$h_\delta(x) = \frac{1}{1+(\frac{1-x}{x})^\delta} \quad \text{if } 0 < x \le 1 \quad \text{and} \quad h_\delta(0) = 0\,.$$

It is immediate that h_1 is the identity function, while if $0 < \delta < 1$ (respectively $\delta > 1$) $h_\delta(x)$ has only one inflection point at $x = 1/2$ and it is concave (resp. convex) on $(0, 1/2)$ and convex (resp. concave) on $(1/2, 1)$.

Notice that h_δ, with $\delta > 1$, penalises the lower and upper tails of a given cdf, while g_β directly discards the outermost $\beta/2$ fraction of points in each tail of the distribution.

3 Distorted M-quantiles and expectiles

Imitating the definition of classical M-quantiles, the *distorted M-quantile* of X of order $r \ge 1$ and level $0 < \alpha < 1$ with distortion function g is defined as a minimizer of $\alpha\mathbb{CE}_{P^*}(X-\theta)_+^r + (1-\alpha)\mathbb{CE}_{\widetilde{P^*}}(X-\theta)_-^r$, that is, as the solution $\theta \in \mathbb{R}$ to the equation

$$\alpha\mathbb{CE}_{P^*}(X-\theta)_+^{r-1} = (1-\alpha)\mathbb{CE}_{\widetilde{P^*}}(X-\theta)_-^{r-1}\,. \tag{4}$$

If $r > 1$, it can be deduced from its alternative representation as the standard M-quantile of some other random variable in Theorem 3.1, that the solution is unique. In the special case that $r = 1$, we take the largest solution to equation (4) in order to fit our purpose of inserting the halfspace regions inside the M-quantile framework.

As for standard M-quantiles, we also omit the order, r, in the notation, and the distorted M-quantile of X of level α (with distortion function g) is denoted as $q_\alpha^*(X)$ in what follows, while the distorted M-quantile associated with the dual distortion function, $\widetilde{g}$ is denoted as $\widetilde{q}_\alpha^*(X)$. Observe that for symmetric distortion functions, such as the trim or the sigmoid ones, both Choquet expectations in (4) are taken with respect to the same distorted probability P^*.

The following result, which is derived using the inverse transform method and the properties of the Choquet expectation, enlightens about the nature and properties of distorted M-quantiles.

Theorem 3.1. *Given two random variables X, Y, two real numbers $r \geq 1$, $0 < \alpha < 1$, and a distortion function g such that $\mathbb{CE}_{P^*} X_+^{r-1}, \mathbb{CE}_{\widetilde{P^*}} X_-^{r-1}, \mathbb{CE}_{P^*} Y_+^{r-1}, \mathbb{CE}_{\widetilde{P^*}} Y_-^{r-1}$ are finite, then*

1. *The distorted M-quantile of X of order r and level α is the M-quantile of order r and level α of a random variable Z (on the same probabiltiy space) whose cdf is $\widetilde{g}(F_X)$, that is,*
$$q_\alpha^*(X) = q_\alpha(Z) \text{ with } F_Z = \widetilde{g}(F_X).$$
2. *Monotonicity on the level. If $0 < \alpha \leq \beta < 1$, then $q_\alpha^*(X) \leq q_\beta^*(X)$.*
3. *Upper and lower M-quantiles. $q_\alpha^*(-X) = -\widetilde{q}_{1-\alpha}^*(X)$ and if g is symmetric, $q_\alpha^*(-X) = -q_{1-\alpha}^*(X)$.*
4. *Translation equivariance. $q_\alpha^*(b + X) = b + q_\alpha^*(X)$ for $b \in \mathbb{R}$.*
5. *Positive homogeneity. $q_\alpha^*(\lambda X) = \lambda q_\alpha^*(X)$ if $\lambda \geq 0$.*
6. *Monotonicity. If $X \leq Y$, then $q_\alpha^*(X) \leq q_\alpha^*(Y)$.*

Proof. In order to prove 1., let Z be a random variable defined on the same probability space of X and whose cdf is $F_Z = \widetilde{g}(F_X)$. Due to the relation $\widetilde{g}(x) = 1 - g(1 - x)$ for all x, the right term of equation (4) can be written in terms of Z as

$$\begin{aligned}
\mathbb{CE}_{P^*}(X - \theta)_+^{r-1} &= \int_0^\infty g\left(P\left((X-\theta)_+^{r-1} \geq x\right)\right) \mathrm{d}x \\
&= \int_0^\infty g\left(1 - F_X(x^{1/(r-1)} + \theta)\right) \mathrm{d}x \\
&= \int_0^\infty 1 - F_Z(x^{1/(r-1)} + \theta) \,\mathrm{d}x \\
&= \mathbb{E}(Z - \theta)_+^{r-1},
\end{aligned}$$

where the last expression appears in terms of an expectation with respect to the probability in the considered probability space, P.

Since we can do a similar transformation with the left term of equation (4), the distorted M-quantile of X of level α matches the undistorted M-quantile of Z of level α.

Statement 2. is immediate after 1. and the monotonicity of the M-quantiles in α, see (Bellini *et al.*, 2014, Prop. 5)

To prove 3., just observe that $\mathbb{CE}_{P^*}(-X - \theta)_+^{r-1} = \mathbb{CE}_{P^*}(X - (-\theta))_-^{r-1}$ and similarly $\mathbb{CE}_{\widetilde{P^*}}(-X - \theta)_-^{r-1} = \mathbb{CE}_{\widetilde{P^*}}(X - (-\theta))_+^{r-1}$, so the weights and distorted probabilities are interchanged between the terms of equation (4) .

Following 1., let $b \in \mathbb{R}$, then

$$\begin{aligned}\mathbb{CE}_{P^*}(X+b-\theta)_+^{r-1} &= \int_0^\infty g\left(P\left((X-(\theta+b))_+^{r-1} \geq x\right)\right)\,\mathrm{d}x \\ &= \int_0^\infty g\left(1-F_X(x^{1/(r-1)}+(\theta-b))\right)\,\mathrm{d}x \\ &= \int_0^\infty 1-F_Z(x^{1/(r-1)}+(\theta-b))\,\mathrm{d}x \\ &= \mathbb{E}(Z+b-\theta)_+^{r-1}.\end{aligned}$$

Similarly, we have that $\mathbb{CE}_{\widetilde{P^*}}(X+b-\theta)_-^{r-1} = \mathbb{E}_P(Z+b-\theta)_-^{r-1}$, so we conclude that $q^*_\alpha(X+b) = q_\alpha(Z+b)$. Since $q_\alpha(Z+b) = q_\alpha(Z)+b$ and $q^*_\alpha(X) = q_\alpha(Z)$ hold, then 4. is proved.

It also holds $\mathbb{CE}_{P^*}(\lambda X-\theta)_+^{r-1} = \mathbb{E}_P(\lambda Z-\theta)_+^{r-1}$ and analogously $\mathbb{CE}_{\widetilde{P^*}}(\lambda X-\theta)_-^{r-1} = \mathbb{E}_P(\lambda Z-\theta)_-^{r-1}$, which proves that $q^*_\alpha(\lambda X) = q_\alpha(\lambda Z)$, and since the undistorted M-quantile satisfies $q_\alpha(\lambda Z) = \lambda q_\alpha(Z)$ then 5. is proved.

If $X \leq Y$ then $(X-\theta)_+^{r-1} \leq (Y-\theta)_+^{r-1}$ hence $\{(X-\theta)_+^{r-1} \geq x\} \subseteq \{(Z-\theta)_+^{r-1} \geq x\}$ and since g is a non-decreasing function, we have

$$\begin{aligned}\mathbb{CE}_{P^*}(X-\theta)_+^{r-1} &= \int_0^\infty g\left(P\left((X-\theta)_+^{r-1} \geq x\right)\right)\,\mathrm{d}x \\ &\leq \int_0^\infty g\left(P\left((Y-\theta)_+^{r-1} \geq x\right)\right)\,\mathrm{d}x \\ &= \mathbb{CE}_{P^*}(Y-\theta)_+^{r-1},\end{aligned}$$

and in the same way that $\mathbb{CE}_{\widetilde{P^*}}(X-\theta)_-^{r-1} \leq \mathbb{CE}_{\widetilde{P^*}}(Y-\theta)_-^{r-1}$, hence $q^*_\alpha(X) \leq q^*_\alpha(Y)$. □

Remark 3.2. The distorted M-quantiles of order $r=2$ (distorted expectiles) are denoted by e^*. The most central distorted expectile (obtained at level $\alpha = 1/2$) of random variable X is the symmetric Choquet expectation of X with respect to $P^* = g(P)$, that is, $e^*_{1/2}(X) = \mathbb{CE}_{P^*}X$, which is the mean of a random variable whose cdf is $\widetilde{g}(F_X)$.

The inverse distorted M-quantile of $x \in \mathbb{R}$ with respect to a random variable X is the level α for which x matches the distorted M-quantile of X, $x = q^*_\alpha(X)$. It is computed solving (4) on α and yields

$$q_X^{*-1}(x) = \left(1 + \frac{\mathbb{CE}_{P^*}(X-x)_+^{r-1}}{\mathbb{CE}_{\widetilde{P^*}}(X-x)_-^{r-1}}\right)^{-1}. \tag{5}$$

We use the symbol $e_X^{*-1}(x)$ to indicate the inverse distorted expectile function, corresponding to (5) for $r=2$.

3.1 Sample distorted M-quantiles

Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ be an ordered sample of real numbers. Let g be a distortion function, and for $i = 1, \ldots, n$ define the weights

$$w_i = \widetilde{g}\left(\frac{i}{n}\right) - \widetilde{g}\left(\frac{i-1}{n}\right) = g\left(1 - \frac{(i-1)}{n}\right) - g\left(1 - \frac{i}{n}\right),$$

hence, it is immediate that $w_i \geq 0$ for all i and $\sum_i w_i = 1$.

For $0 < \alpha < 1$, as in (4), the *sample distorted M-quantile*, denoted $q^*_{n,\alpha}(x_{(1)}, \ldots, x_{(n)})$ or simply $q^*_{n,\alpha}$ when there is no possible confusion with the sample, can be computed as the solution to the equation

$$\alpha \sum_{x_{(i)} \geq q^*_{n,\alpha}} w_i (x_{(i)} - q^*_{n,\alpha})^{r-1} = (1-\alpha) \sum_{x_{(i)} < q^*_{n,\alpha}} w_i (q^*_{n,\alpha} - x_{(i)})^{r-1} . \tag{6}$$

Notice from (6) that the corresponding sample inverse distorted M-quantile is

$$q_n^{*-1}(x) = \left(1 + \frac{\displaystyle\sum_{x_{(i)} > x} w_i (x_{(i)} - x)^{r-1}}{\displaystyle\sum_{x_{(i)} < x} w_i (x - x_{(i)})^{r-1}}\right)^{-1} . \tag{7}$$

Alike the sample expectile, the sample distorted expectile ($r = 2$), denoted by $e^*_{n,\alpha}$ can be calculated from (6) by means of *iterated re-weighting*, while its inverse follows (7) and is denoted by $e^{*-1}_{n,\alpha}$.

4 Distorted M-quantile central regions and depth

4.1 Distorted M-quantile central regions

A *central region* associated with a random vector contains valuable information about the location, scatter, and dependency structure among the components in its location, size, and shape. Based on the (univariate) distorted M-quantiles, we build the distorted M-quantile central regions in terms of an intersection of halfspaces.

Consider a d-dimensional random vector $\boldsymbol{X}$ such that $\mathbb{CE}_{P^*}\|\boldsymbol{X}\|^{r-1}$ and $\mathbb{CE}_{\widetilde{P^*}}\|\boldsymbol{X}\|^{r-1}$ are finite for some $r \geq 1$ and $0 < \alpha \leq 1$. The *distorted M-quantile region* of order r and level α is defined as

$$\mathrm{Q}^*_\alpha(\boldsymbol{X}) = \bigcap_{\boldsymbol{u} \in \mathbb{S}^{d-1}} \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{u} \rangle \leq q^*_{1-\alpha}(\langle \boldsymbol{X}, \boldsymbol{u} \rangle)\} , \tag{8}$$

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

Notice that in the univariate case ($d=1$), the region is the closed interval

$$\mathrm{Q}^*_\alpha(X) = [-q^*_{1-\alpha}(-X), q^*_{1-\alpha}(X)] = [\widetilde{q}^*_\alpha(X), q^*_{1-\alpha}(X)].$$

If the distortion function g is the identity function, the (non distorted) M-quantile regions obtained were already studied in Daouia and Paindaveine (2019), while if further $r=2$, the expectile regions were the specific focus of Cascos and Ochoa (2021).

From its construction and Theorem 3.1, it is clear that the distorted M-quantile regions satisfy the properties that are stated below.

Proposition 4.1. *Let $\boldsymbol{X}$ be a d-dimensional random vector, fix $0<\alpha\le 1$, and $r\ge 1$ such that $\mathbb{CE}_{P^*}\|\boldsymbol{X}\|^{r-1}$ and $\mathbb{CE}_{\widetilde{P^*}}\|\boldsymbol{X}\|^{r-1}$ are finite, then*

1. Compactness and convexity. *$\mathrm{Q}^*_\alpha(\boldsymbol{X})$ is a compact and convex subset of $\mathbb{R}^d$.*

2. Nesting. *The lower the level, the wider the region: $\mathrm{Q}^*_\alpha(\boldsymbol{X})\subseteq \mathrm{Q}^*_\beta(\boldsymbol{X})$ if $0<\beta\le\alpha\le 1$.*

3. Affine equivariance. *$\mathrm{Q}^*_\alpha(A\boldsymbol{X}+\boldsymbol{b}) = A\mathrm{Q}^*_\alpha(\boldsymbol{X})+\boldsymbol{b}$ for any nonsingular matrix $A\in\mathbb{R}^{d\times d}$ and $\boldsymbol{b}\in\mathbb{R}^d$.*

4. Contains the centermost point. *If the distribution of $\boldsymbol{X}$ is centrally symmetric about $\boldsymbol{y}_0$, that is, $P(\boldsymbol{y}_0-\boldsymbol{X}\in H) = P(\boldsymbol{X}-\boldsymbol{y}_0\in H)$ for every halfspace $H\subseteq\mathbb{R}^d$, then $\boldsymbol{y}_0\in\mathrm{Q}^*_\alpha(\boldsymbol{X})$ whenever $\mathrm{Q}^*_\alpha(\boldsymbol{X})$ is nonvoid.*

Proof. Notice that the compactness and the convexity of these regions is due to way that they are obtained as intersection of closed halfspaces, the nesting is immediate after the monotonicity of the M-quantiles in α, whereas the affine equivariance is a direct consequence of the translation equivariance and positive homogeneity of distorted M-quantiles, presented in Theorem 3.1.

We will only show here that for centrally symmetric distributions, every nonvoid distorted M-quantile central region contains the point of symmetry. Denote the point of symmetry by $\boldsymbol{y}_0$ and observe that the central region $\mathrm{Q}^*_\alpha(\boldsymbol{X})$ introduced in (8) can be rewritten as

$$\boldsymbol{y}_0+\mathrm{Q}^*_\alpha(\boldsymbol{X}-\boldsymbol{y}_0) = \bigcap_{\boldsymbol{u}\in\mathbb{S}^{d-1}}\{\boldsymbol{y}_0+\boldsymbol{x}\in\mathbb{R}^d : -q^*_{1-\alpha}(\langle\boldsymbol{y}_0-\boldsymbol{X},\boldsymbol{u}\rangle)\le\langle\boldsymbol{x},\boldsymbol{u}\rangle\le q^*_{1-\alpha}(\langle\boldsymbol{X}-\boldsymbol{y}_0,\boldsymbol{u}\rangle)\}.$$

After (Zuo and Serfling, 2000c, Lem. 2.1), the distribution of $\boldsymbol{X}$ is centrally symmetric about $\boldsymbol{y}_0$ if and only if for every $\boldsymbol{u}\in\mathbb{S}^{d-1}$, the random variables $\langle\boldsymbol{X}-\boldsymbol{y}_0,\boldsymbol{u}\rangle$ and $\langle\boldsymbol{y}_0-\boldsymbol{X},\boldsymbol{u}\rangle$ share the same distribution, so their distorted M-quantiles must be identical. Consequently, either $q^*_{1-\alpha}(\langle\boldsymbol{X}-\boldsymbol{y}_0,\boldsymbol{u}\rangle)\ge 0$ for all $\boldsymbol{u}\in\mathbb{S}^{d-1}$ and then $\boldsymbol{y}_0\in\mathrm{Q}^*_\alpha(\boldsymbol{X})$, or alternatively $\mathrm{Q}^*_\alpha(\boldsymbol{X})$ is void. □

4.2 Distorted M-quantile sample regions

Given a sample $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ and a unitary vector $\boldsymbol{u} \in \mathbb{S}^{d-1}$, let $\pi_{\boldsymbol{u}}$ be the permutation of $\{1, 2, \ldots, n\}$ such that $\langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(1)}, \boldsymbol{u}\rangle \leq \ldots \leq \langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(n)}, \boldsymbol{u}\rangle$, then for $0 < \alpha \leq 1$, the sample distorted M-quantile region is either void or, for sufficiently small α, yields the polyhedral set:

$$\mathrm{Q}^*_{n,\alpha} = \bigcap_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \{\boldsymbol{x} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{u}\rangle \leq q^*_{n,1-\alpha}(\langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(1)}, \boldsymbol{u}\rangle, \ldots, \langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(n)}, \boldsymbol{u}\rangle)\}\,.$$

In Figure 1 we illustrate the robustness to outliers of the distorted M-quantile regions with the trim distortion function in comparison with the undistorted ones. This illustration includes the expectile regions and the halfspace ones.

For this example we have simulated 190 observations of a bivariate normal random vector with mean vector $(10, 20)^\top$ and covariance matrix $\begin{pmatrix} 5 & 4 \\ 4 & 4 \end{pmatrix}$ and 10 further observations of a bivariate normal random vector with mean $(16, 16)^\top$ and identity covariance matrix. Specifically, Figure 1 presents, on the left, several distorted M-quantile regions (contours of the halfspace regions, $r = 1$, at the top, of the M-quantile regions for $r = 1.5$ in the second row, of the expectile regions, $r = 2$, in the third row, and contours of the M-quantile regions for $r = 3$ at the bottom) with a trim distortion level $\beta = 0.1$ and, on the right, the corresponding (undistorted) M-quantile regions. The depth levels considered are 0.005, 0.02, 0.05, 0.1, 0.2, 0.3, and 0.4.

Observe that the outermost halfspace regions (top right) are highly affected in the presence of outliers, while their trimmed counterparts (top left) barely change in shape with respect to the inner regions. These inner regions are almost identical on both sides. The situation with the remaining outer M-quantile regions is almost the same, but changes substantially with the inner regions. The greater the level r is, the more the M-quantile inner regions are also affected by the outliers. This situation is corrected in the distorted M-quantile regions.

4.3 Distorted M-quantile data depth function

Once we have introduced the M-quantile regions, it is possible to construct the corresponding *distorted M-quantile depth* for any fixed order $r \geq 1$ in a natural way. The depth of each point $\boldsymbol{y} \in \mathbb{R}^d$ is the maximum level α such that it belongs to the region $\mathrm{Q}^*_\alpha(\boldsymbol{X})$, that is,

$$\mathrm{QD}^*(\boldsymbol{y}; \boldsymbol{X}) = \sup\{0 < \alpha \leq 1 : \boldsymbol{y} \in \mathrm{Q}^*_\alpha(\boldsymbol{X})\}. \tag{9}$$

The so-defined distorted M-quantile depth satisfies the usual properties of a depth function.

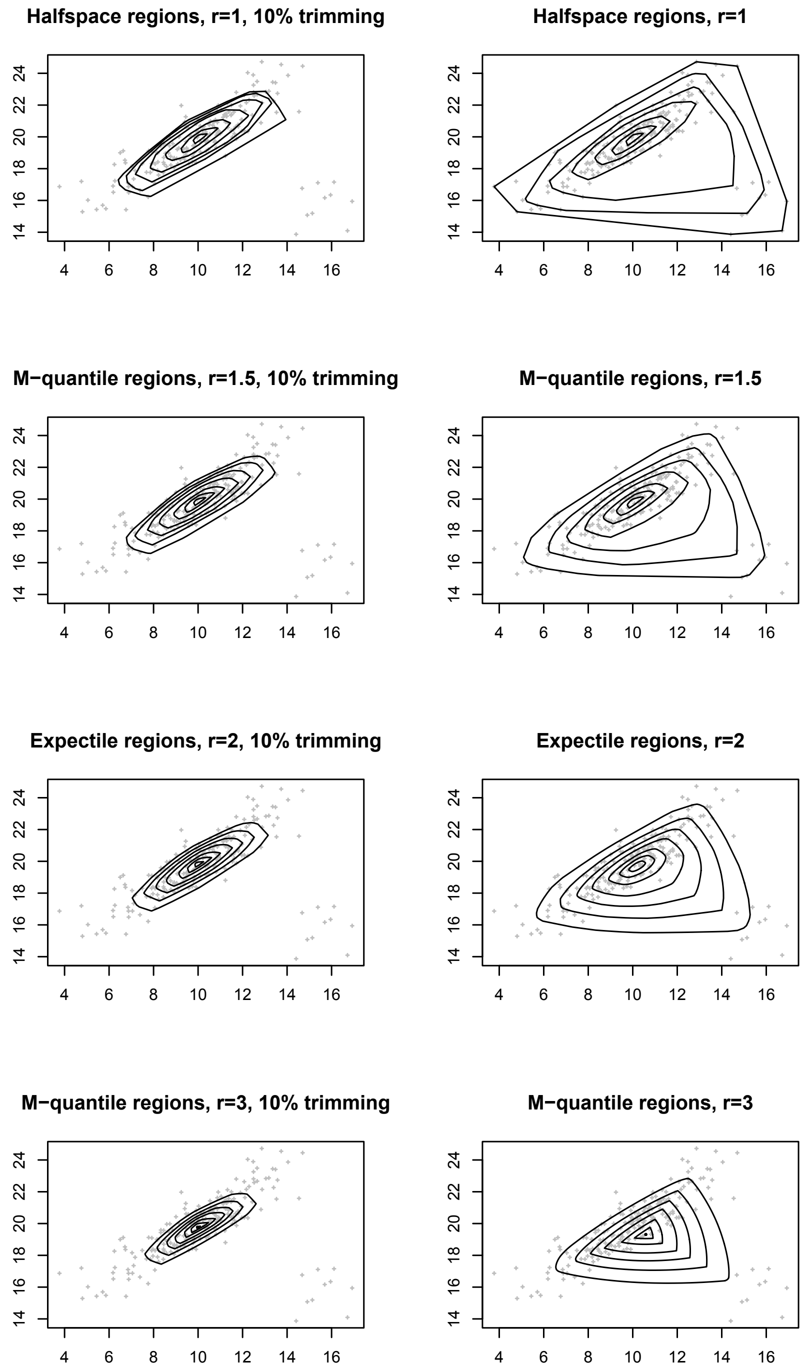

Figure 1: M-quantile regions including halfspace and expectile ones with 10% of trimming and undistorted.

Proposition 4.2. *Let $\boldsymbol{X}$ be a d-dimensional random vector and fix $r \geq 1$ such that $\mathbb{CE}_{P^*}\|\boldsymbol{X}\|^{r-1}$ and $\mathbb{CE}_{\widetilde{P^*}}\|\boldsymbol{X}\|^{r-1}$ are finite, then*

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

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

2. Upper semicontinuity. $\mathrm{QD}^*(\,\cdot\,; \boldsymbol{X})$ *is an upper semicontinuous function.*

3. Maximality at center. *If the distribution of $\boldsymbol{X}$ is centrally symmetric about $\boldsymbol{y}_0$, then* $\mathrm{QD}^*(\,\cdot\,; \boldsymbol{X})$ *attains its maximum value at $\boldsymbol{y}_0$.*

4. Decreasing on rays. *If $\boldsymbol{y}_0 \in \mathbb{R}^d$ is the deepest point and $\boldsymbol{y} \in \mathbb{R}^d$, then for $0 \leq \lambda \leq 1$ it holds*

$$\mathrm{QD}^*(\boldsymbol{y}; \boldsymbol{X}) \leq \mathrm{QD}^*(\lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{y}_0; \boldsymbol{X}).$$

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

It is not hard to realize that these properties follow from the ones of the distorted M-quantile regions. Notice further that, unlike for the (undistorted) expectile depth, the maximum value of the distorted expectile depth of a given distribution is not necessarily 1/2, while, in the univariate setting, the distorted M-quantile depth renders

$$\mathrm{QD}^*(y; X) = \min\{1 - q_X^{*-1}(y), 1 - q_{-X}^{*-1}(-y)\}.$$

The result below gives us a useful expression to compute the multivariate distorted M-quantile depth of a given point.

Theorem 4.3. *The distorted M-quantile depth of a point $\boldsymbol{y} \in \mathbb{R}^d$ with respect to the distribution of the d-dimensional random vector $\boldsymbol{X}$ satisfies:*

$$\mathrm{QD}^*(\boldsymbol{y}; \boldsymbol{X}) = \left(1 + \sup_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\mathbb{CE}_{\widetilde{P^*}}\langle \boldsymbol{X} - \boldsymbol{y}, \boldsymbol{u}\rangle_-^{r-1}}{\mathbb{CE}_{P^*}\langle \boldsymbol{X} - \boldsymbol{y}, \boldsymbol{u}\rangle_+^{r-1}}\right)^{-1}.$$

Proof. Just notice that

$$\begin{aligned}
\mathrm{QD}^*(\boldsymbol{y};\boldsymbol{X}) &= \sup\{0<\alpha\le 1:\, \boldsymbol{y}\in \mathrm{Q^*}_{\alpha}(\boldsymbol{X})\}\\
&= \sup\left\{0<\alpha\le 1:\, \langle \boldsymbol{y},\boldsymbol{u}\rangle \le q^*_{1-\alpha}(\langle \boldsymbol{X},\boldsymbol{u}\rangle) \quad \text{for all } \boldsymbol{u}\in\mathbb{S}^{d-1}\right\}\\
&= \sup\left\{0<\alpha\le 1:\, q^{*-1}_{\langle \boldsymbol{X},\boldsymbol{u}\rangle}(\langle \boldsymbol{y},\boldsymbol{u}\rangle) \le 1-\alpha \quad \text{for all } \boldsymbol{u}\in\mathbb{S}^{d-1}\right\}\\
&= \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \left(1-q^{*-1}_{\langle \boldsymbol{X},\boldsymbol{u}\rangle}(\langle \boldsymbol{y},\boldsymbol{u}\rangle)\right)\\
&= \left(1+\sup_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\mathbb{CE}_{\widetilde{P^*}}\langle \boldsymbol{X}-\boldsymbol{y},\boldsymbol{u}\rangle_-^{r-1}}{\mathbb{CE}_{P^*}\langle \boldsymbol{X}-\boldsymbol{y},\boldsymbol{u}\rangle_+^{r-1}}\right)^{-1}
\end{aligned}$$

because of (5). □

The special case of Theorem 4.3 for $r = 2$ is used in Appendix A to build an algorithm for the exact computation of the bivariate distorted expectile depth.

Observe that if the distortion function g is symmetric then the distorted M-quantile depth yields the very appealing expression

$$\mathrm{QD}^*(\boldsymbol{y};\boldsymbol{X}) = \inf_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \frac{\mathbb{CE}_{P^*}\langle \boldsymbol{X}-\boldsymbol{y},\boldsymbol{u}\rangle_+^{r-1}}{\mathbb{CE}_{P^*}|\langle \boldsymbol{X}-\boldsymbol{y},\boldsymbol{u}\rangle|^{r-1}}\,.$$

4.4 Multivariate ranks based on the distorted M-quantile depth

Following the ideas summarized in Cascos (2010) and given a multivariate dataset, we can assess ranks to the points in it in terms of their distorted M-quantile depths with respect to the dataset. After computing the depth of every point, rank 1 is assessed to the k points with the lowest depth, rank $k+1$ to the points with the second lowest depth, and so on, until the highest rank is assessed to the deepest point of the dataset. By means of this rank, we have a notion of ordering in $\mathbb{R}^d$ which could be used in further statistical analysis, see for example Liu and Singh (1992).

In Figure 2 we have plotted 39 points simulated from a bivariate normal distribution centered at the origin of coordinates $(0,0)^\top$ with covariance matrix $\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$ along with an extreme outlier at $(7,-7)^\top$ and ranked them by means of several distorted M-quantile depth functions: the halfspace one (top left), the expectile (top right), the distorted expectile with the trim distortion function $g_{0.1}$ (10% of trimming) and the sigmoid distortion with parameter $\delta = 3$. Notice that for the expectile depth, the highest ranks are awarded to the points closest to the outlier, among those that were simulated with regard to the normal distribution with mean at the origin of coordinates. When some observations are trimmed, the highest rank moves toward the origin of coordinates, but some inner points also assume rank 1. Finally, when the sigmoid distortion function is used, the points with the highest ranks are close to the origin of coordinates while all points inside the interior of the convex hull of the data cloud assume a rank greater than 1.

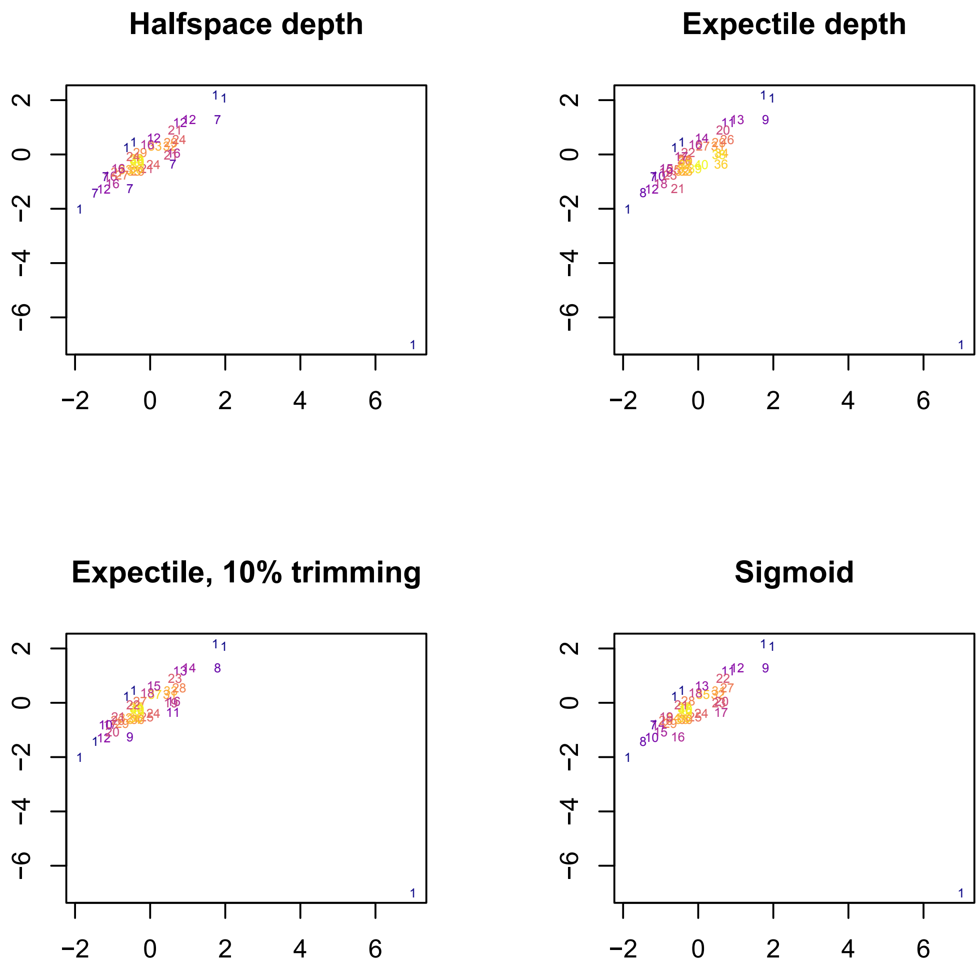

Figure 2: Depth-based ranks for a bivariate normal sample with an outlier with various distorted M-quantile depths.

The halfspace depth was calculated using the R package `ddalpha` by Pokotylo *et al.* (2019).

5 Distorted M-quantile depth and regression models

Assume that the response of a linear regression model is given in terms of a d-dimensional random vector $\boldsymbol{Y}$, while there are p regressors forming a p-dimensional random vector $\boldsymbol{X}$.

If $\boldsymbol{x} \in \mathbb{R}^p$, we use the notation $\mathring{\boldsymbol{x}}$ to represent the element in $\mathbb{R}^{p+1}$ whose first component is 1, and the remaining components are equal to those of $\boldsymbol{x}$. For each observation $i = 1, \ldots, n$, we can write the multivariate regression model as

$$\boldsymbol{Y}_i = \Theta \mathring{\boldsymbol{X}}_i + \boldsymbol{\epsilon}_i \,,$$

where Θ is a $(p+1) \times d$ matrix with each row formed by the regression coefficients of the multiple linear regression model that fits each of the components of random vector $\boldsymbol{Y}$, and $\boldsymbol{\epsilon}_i$ is the (vector) error term associated with the i-th observation.

Instead of estimating matrix Θ, we will focus on studying the conditional central regions of $\boldsymbol{Y}$ for any level α and given the value assumed by the explanatory vector $\boldsymbol{X}$.

5.1 Distorted M-quantile regression hyperplanes for a univariate response

By analogy with (1), for a univariate response Y and a p-dimensional vector of regressors $\boldsymbol{X}$, the *distorted M-quantile regression hyperplane of order* $r \geq 1$ *and level* $0 < \alpha \leq 1$ is defined through the vector of coefficients $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_p) \in \mathbb{R}^{p+1}$ minimizing the expression

$$\boldsymbol{q}^*_\alpha(Y|\boldsymbol{X}) = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^{p+1}} \left[\alpha \mathbb{C}\mathbb{E}_{P^*}\left[Y - \langle \mathring{\boldsymbol{X}}, \boldsymbol{\theta}\rangle\right]_+^r + (1-\alpha)\mathbb{C}\mathbb{E}_{\widetilde{P^*}}\left[Y - \langle \mathring{\boldsymbol{X}}, \boldsymbol{\theta}\rangle\right]_-^r\right] \tag{10}$$

and results to be $H^*_\alpha(Y|\boldsymbol{X}) = \{(\boldsymbol{x}, y) \in \mathbb{R}^{p+1} : \langle \boldsymbol{q}^*_\alpha(Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}\rangle = y\}$. When the distortion function is $g(x) = x$, we denote the solution as $\boldsymbol{q}_\alpha(Y|\boldsymbol{X})$ and the M-quantile regression hyperplane of order $r \geq 1$ as $H_\alpha(Y|\boldsymbol{X})$, while notations $\widetilde{\boldsymbol{q}}^*_\alpha(Y|\boldsymbol{X})$ and $\widetilde{H}^*_\alpha(Y|\boldsymbol{X})$ correspond to the coefficients and hyperplane obtained for the dual distortion function.

The coefficients of the distorted M-quantile regression hyperplanes fulfill properties similar to those presented in Theorem 3.1 for the distorted M-quantiles.

Proposition 5.1. *For a response random variable Y, $r \geq 1$, $0 < \alpha \leq 1$, and a p-variate vector of regressors $\boldsymbol{X}$, it holds*

1. *Upper and lower M-quantiles.* $\boldsymbol{q}^*_\alpha(-Y|\boldsymbol{X}) = -\widetilde{\boldsymbol{q}}^*_{1-\alpha}(Y|\boldsymbol{X})$ *and if the distortion g is symmetric,* $\boldsymbol{q}^*_\alpha(-Y|\boldsymbol{X}) = -\boldsymbol{q}^*_{1-\alpha}(Y|\boldsymbol{X})$.
2. *Translation equivariance.* $\boldsymbol{q}^*_\alpha(b + Y|\boldsymbol{X}) = (b, 0, \ldots, 0) + \boldsymbol{q}^*_\alpha(Y|\boldsymbol{X})$ *for any* $b \in \mathbb{R}$.
3. *Positive homogeneity.* $\boldsymbol{q}^*_\alpha(\lambda Y|\boldsymbol{X}) = \lambda \boldsymbol{q}^*_\alpha(Y|\boldsymbol{X})$ *if* $\lambda \geq 0$.

In terms of the M-quantile regression hyperplane, 1. means that $H^*_\alpha(-Y|\boldsymbol{X})$ is identical to $\widetilde{H}^*_{1-\alpha}(Y|\boldsymbol{X})$ except for a change of sign in the last coordinate; 2. states that the only coefficient affected by a translation in Y is the one corresponding to the offset of the hyperplane, θ_0, so $H^*_\alpha(b+Y|\boldsymbol{X})$ matches $H^*_\alpha(Y|\boldsymbol{X})$ with a translation of b units in the last coordinate; finally 3. states that when the response is rescaled then $H^*_\alpha(\lambda Y|\boldsymbol{X})$ matches $H^*_\alpha(Y|\boldsymbol{X})$ except for the fact that the last coordinate is multiplied by λ.

Due to the similarities in the way $q^*_\alpha(Y)$ and $\boldsymbol{q}^*_\alpha(Y|\boldsymbol{X})$ are defined, the proof of Proposition 5.1 is similar to the proof of Theorem 3.1.

5.2 Distorted M-quantile regression regions and conditional distorted M-quantile regions

We use the name regression regions to refer to sets in the $(p+d)$-dimensional space containing the observations of both, the explanatory and response variables, while under conditional regions given some specific value of the explanatory variables, we refer to sets in the d-dimensional space of the response variables.

5.2.1 Univariate response

The *distorted M-quantile regression region of level α* in a simple linear regression model is the set of points in $\mathbb{R}^{p+1}$ comprised between the regression hyperplanes $H^*_{1-\alpha}(Y|\boldsymbol{X})$ and $\widetilde{H}^*_\alpha(Y|\boldsymbol{X})$ whenever $H^*_{1-\alpha}(Y|\boldsymbol{X})$ lies above $\widetilde{H}^*_\alpha(Y|\boldsymbol{X})$ on the last coordinate (the one associated with the response variable). Since $-\boldsymbol{q}^*_{1-\alpha}(-Y|\boldsymbol{X}) = \widetilde{\boldsymbol{q}}^*_\alpha(Y|\boldsymbol{X})$, see Proposition 5.1, we can write it as

$$\mathrm{Q}^*_\alpha(Y|\boldsymbol{X}) = \left\{(\boldsymbol{x},y)\in\mathbb{R}^{p+1} :\ \langle -\boldsymbol{q}^*_{1-\alpha}(-Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}\rangle \le y \le \langle \boldsymbol{q}^*_{1-\alpha}(Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}\rangle\right\},$$

and recall that if g is symmetric, $-\boldsymbol{q}^*_{1-\alpha}(-Y|\boldsymbol{X}) = \boldsymbol{q}^*_\alpha(Y|\boldsymbol{X})$.

The *conditional distorted M-quantile region when $\boldsymbol{X}=\boldsymbol{x}_0$* is the interval given by the projection on the last coordinate of the intersection of $\mathrm{Q}^*_\alpha(Y|\boldsymbol{X})$ with the affine line constituted by all points whose first p coordinates are identical to $\boldsymbol{x}_0$. We can indistinctively write this region using set or interval notation,

$$\begin{aligned}\mathrm{Q}^*_\alpha(Y|\boldsymbol{X}=\boldsymbol{x}_0) &= \{y\in\mathbb{R} :\ (\boldsymbol{x}_0,y)\in \mathrm{Q}^*_\alpha(Y|\boldsymbol{X})\}\\ &= \left[\langle -\boldsymbol{q}^*_{1-\alpha}(-Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}_0\rangle, \langle \boldsymbol{q}^*_{1-\alpha}(Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}_0\rangle\right],\end{aligned}$$

and it is void for some values of α and $\boldsymbol{x}_0$. If g is symmetric, the interval adopts the form $\left[\langle \boldsymbol{q}^*_\alpha(Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}_0\rangle, \langle \boldsymbol{q}^*_{1-\alpha}(Y|\boldsymbol{X}), \mathring{\boldsymbol{x}}_0\rangle\right]$.

After Proposition 5.1, it is immediate that the univariate conditional distorted M-quantile regions are affine equivariant, since $\mathrm{Q}^*_\alpha(b+aY|\boldsymbol{X}=\boldsymbol{x}_0) = b + a\mathrm{Q}^*_\alpha(Y|\boldsymbol{X}=\boldsymbol{x}_0)$ for $a,b\in\mathbb{R}$.

5.2.2 General (multivariate) response

If the response $\boldsymbol{Y}$ is a d-dimensional random vector, we define the *distorted M-quantile regression regions* in terms of intersections of halfspaces. Specifically, all univariate projections of the response are obtained after multiplication times every element in the unit sphere $\mathbb{S}^{d-1}$. The regression hyperplane of all the projections of the response are obtained, and finally the regression regions are produced as intersections of halfspaces whose boundaries are precisely the regression hyperplanes,

$$\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X}) = \bigcap_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \left\{(\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}^{p+d}:\ \langle\boldsymbol{y},\boldsymbol{u}\rangle \le \langle \boldsymbol{q}^*_{1-\alpha}(\langle\boldsymbol{Y},\boldsymbol{u}\rangle|\boldsymbol{X}),\mathring{\boldsymbol{x}}\rangle\right\}.$$

The *distorted M-quantile conditional region* of level α and given that the regressors assume value $\boldsymbol{x}_0\in\mathbb{R}^p$ is the subset of $\mathbb{R}^d$ obtained as the projection on the last d coordinates of the intersection of the distorted M-quantile regression region $\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X})$ with the d-dimensional affine space constituted by all points whose first p coordinates are identical to $\boldsymbol{x}_0$,

$$\begin{aligned}\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_0) =& \{\boldsymbol{y}\in\mathbb{R}^d:\ (\boldsymbol{x}_0,\boldsymbol{y})\in\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X})\}\\ =& \bigcap_{\boldsymbol{u}\in\mathbb{S}^{d-1}}\left\{\boldsymbol{y}\in\mathbb{R}^d:\ \langle\boldsymbol{y},\boldsymbol{u}\rangle\le\langle\boldsymbol{q}^*_{1-\alpha}(\langle\boldsymbol{Y},\boldsymbol{u}\rangle|\boldsymbol{X}),\mathring{\boldsymbol{x}}_0\rangle\right\}.\end{aligned}$$

Despite the fact that distorted M-quantile conditional regions might be void for certain values of α and $\boldsymbol{x}_0$, they still satisfy some of the usual properties of the classical central regions:

1. *Compactness and convexity.* $\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_0)$ is a compact and convex set in $\mathbb{R}^d$.
2. *Affine invariance.* $\mathrm{Q}^*_\alpha(A\boldsymbol{Y}+\boldsymbol{b}|\boldsymbol{X}=\boldsymbol{x}_0) = A\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_0)+\boldsymbol{b}$ for any nonsingular matrix $A\in\mathbb{R}^{d\times d}$ and vector $\boldsymbol{b}\in\mathbb{R}^d$.

It is immediate that these properties are derived from the ones of the M-quantile regression hyperplanes. Observe that, among the properties of the conditional regions, there is not a nesting one. The reason is that the regression hyperplanes at different levels might well intersect and consequently one regression or conditional region of some given level might contain some points that do not lie inside the corresponding region of a lower level.

5.3 Conditional distorted M-quantile depth

Once the conditional regions have been defined, the *conditional distorted M-quantile depth of a point $\boldsymbol{y}\in\mathbb{R}^d$ given that the regressors assume value $\boldsymbol{x}_0\in\mathbb{R}^p$* is obtained, as usual, by

$$\mathrm{QD}^*(\boldsymbol{y};\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_0) = \sup\{0<\alpha\le 1:\ \boldsymbol{y}\in\mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_0)\}.$$

Notice that this depth can also be defined in terms of the distorted M-quantile regression regions as the supremum of all levels α such that $(\boldsymbol{x}_0, \boldsymbol{y}) \in \mathrm{Q}^*_\alpha(\boldsymbol{Y}|\boldsymbol{X})$.

Intuitively, the so-defined conditional distorted M-quantile regression depth satisfies some of the usual properties of a depth function:

1. *Affine invariance.* The conditional depth of a point $\boldsymbol{y} \in \mathbb{R}^d$ is independent of the underlying coordinate system in the space of the response variables, that is, for any nonsingular matrix $A \in \mathbb{R}^{d\times d}$ and $\boldsymbol{b} \in \mathbb{R}^d$, it holds $\mathrm{QD}^*(A\boldsymbol{y} + \boldsymbol{b}; A\boldsymbol{Y} + \boldsymbol{b}|\boldsymbol{X} = \boldsymbol{x}_0) = \mathrm{QD}^*(\boldsymbol{y}; \boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}_0)$.

2. *Upper semicontinuity.* The function $\mathrm{QD}^*(\,\cdot\,; \boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}_0)$ is upper semicontinuous.

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

5.4 Sample distorted M-quantile regression regions and depth

Consider a sample of n joint observations of a d-dimensional response variable $\boldsymbol{y}_i$ and a p-dimensional predictor $\boldsymbol{x}_i$ for $1 \le i \le n$, let $\boldsymbol{u} \in \mathbb{S}^{d-1}$ be any unit vector and denote by $\pi_{\boldsymbol{u}} : \{1, \dots, n\} \to \{1, \dots, n\}$ the permutation that sorts the sample points in such a way that the univariate projections of the response variables are ordered in an increasing manner, $\langle \boldsymbol{y}_{\pi_{\boldsymbol{u}(1)}}, \boldsymbol{u}\rangle \le \cdots \le \langle \boldsymbol{y}_{\pi_{\boldsymbol{u}(n)}}, \boldsymbol{u}\rangle$. Given $0 < \alpha \le 1$, the sample distorted M-quantile hyperplane of order r in direction $\boldsymbol{u}$ is determined by the coefficients $\boldsymbol{\theta} = (\theta_0, \dots, \theta_p)$ that minimize the expression

$$\arg\min_{\boldsymbol{\theta}\in\mathbb{R}^{p+1}} \alpha \sum_{i=1}^{n} w_i \left[\langle \boldsymbol{y}_{\pi_{\boldsymbol{u}(i)}}, \boldsymbol{u}\rangle - \langle \mathring{\boldsymbol{x}}_{\pi_{\boldsymbol{u}(i)}}, \boldsymbol{\theta}\rangle\right]_+^r + (1-\alpha) \sum_{i=1}^{n} w_i \left[\langle \boldsymbol{y}_{\pi_{\boldsymbol{u}(i)}}, \boldsymbol{u}\rangle - \langle \mathring{\boldsymbol{x}}_{\pi_{\boldsymbol{u}(i)}}, \boldsymbol{\theta}\rangle\right]_-^r , \tag{11}$$

which are denoted by $\boldsymbol{q}^*_{n,\alpha}(\boldsymbol{u})$. As usual, $w_i = \widetilde{g}\,(i/n) - \widetilde{g}\,((i-1)/n) = g\,(1-(i-1)/n) - g\,(1 - i/n)$ for all i.

The sample conditional (distorted) M-quantile region at level α given $\boldsymbol{x}_0 \in \mathbb{R}^p$ is obtained as

$$\mathrm{Q}^*_{n,\alpha}(\boldsymbol{x}_0) = \bigcap_{\boldsymbol{u}\in\mathbb{S}^{d-1}} \left\{\boldsymbol{y} \in \mathbb{R}^d : \langle \boldsymbol{y}, \boldsymbol{u}\rangle \le \langle \boldsymbol{q}^*_{n,1-\alpha}(\boldsymbol{u}), \mathring{\boldsymbol{x}}_0\rangle\right\} ,$$

while the depth of $\boldsymbol{y} \in \mathbb{R}^d$ given $\boldsymbol{x}_0$ is

$$\mathrm{QD}^*_n(\boldsymbol{y}\,|\,\boldsymbol{x}_0) = \sup\{0 < \alpha \le 1 : \boldsymbol{y} \in \mathrm{Q}^*_{n,\alpha}(\mathring{\boldsymbol{x}}_0)\} .$$

We elude the regression regions in this empirical section, since their basic goal is to serve in the construction of the conditional regions, but they can be simply introduced in the same manner of their population counterparts.

In Figure 3 we have represented the average of the *educational index* (Y_1), the *life expectancy* (Y_2) and the *income index* (X) as defined in `http://hdr.undp.org/en/data`. This database contains information about human development from countries all over the world with records ranging between 1990 and 2019. The conditional regression regions were built given the 0.05, 0.5, and 0.95 quantiles of X by considering M-quantile regression models with several power loss functions, including $r = 1$ (halfspace depth regions) and $r = 2$ (expectile regions). The levels α are taken as 10 equispaced values in interval $[0, 0.45]$. Notice that some of those regions happen to be void. The identity distortion function (undistorted regions) has been used to obtain the charts on the left, while the charts on the right were built using $g_{0.1}$ as distortion function (10% of trimming). At first sight, all the charts suggest that high values of indexes Y_1 and Y_2 are associated with high values of the income index X, which is more evident for the 0.05 and 0.95 quantiles of X. Notice that the conditional regions capture the variability of the data. Interestingly, it seems that the higher the value of the regressor, the lower the variability of the conditional distribution of the response variables, which is particularly evident when considering the trimming level of 10%.

6 Conclusions

After a short introduction to the notion of distorted M-quantiles with power loss functions, emphasizing the distorted expectiles, some novel notions of central regions and data depth are proposed. Both, the distorted M-quantile depth and regions, meet the classical properties expected from any depth function and central regions, that is, the distorted M-quantile depth function is affine invariant, attains a unique maximum value at the center of the distribution (whenever such a center exists), decreases through rays from the deepest point, and vanishes at infinity, whereas their associated regions form a family of compact, convex and nested sets which are equivariant under affine transformations. The sample distorted M-quantile depth and regions were introduced and some algorithm is provided for their computation in the bivariate case.

The distorted M-quantile conditional regression regions, obtained through a multiple output M-quantile regression model, seem to capture relevant information of the joint distribution of the explained variables conditioned to a given value of the regressors. Henceforth, we consider that the conditional regression regions, together with their associated regression depth, could be useful as a multivariate data analysis tool. In the future, we plan to design specific algorithms for the computation of these regression depths.

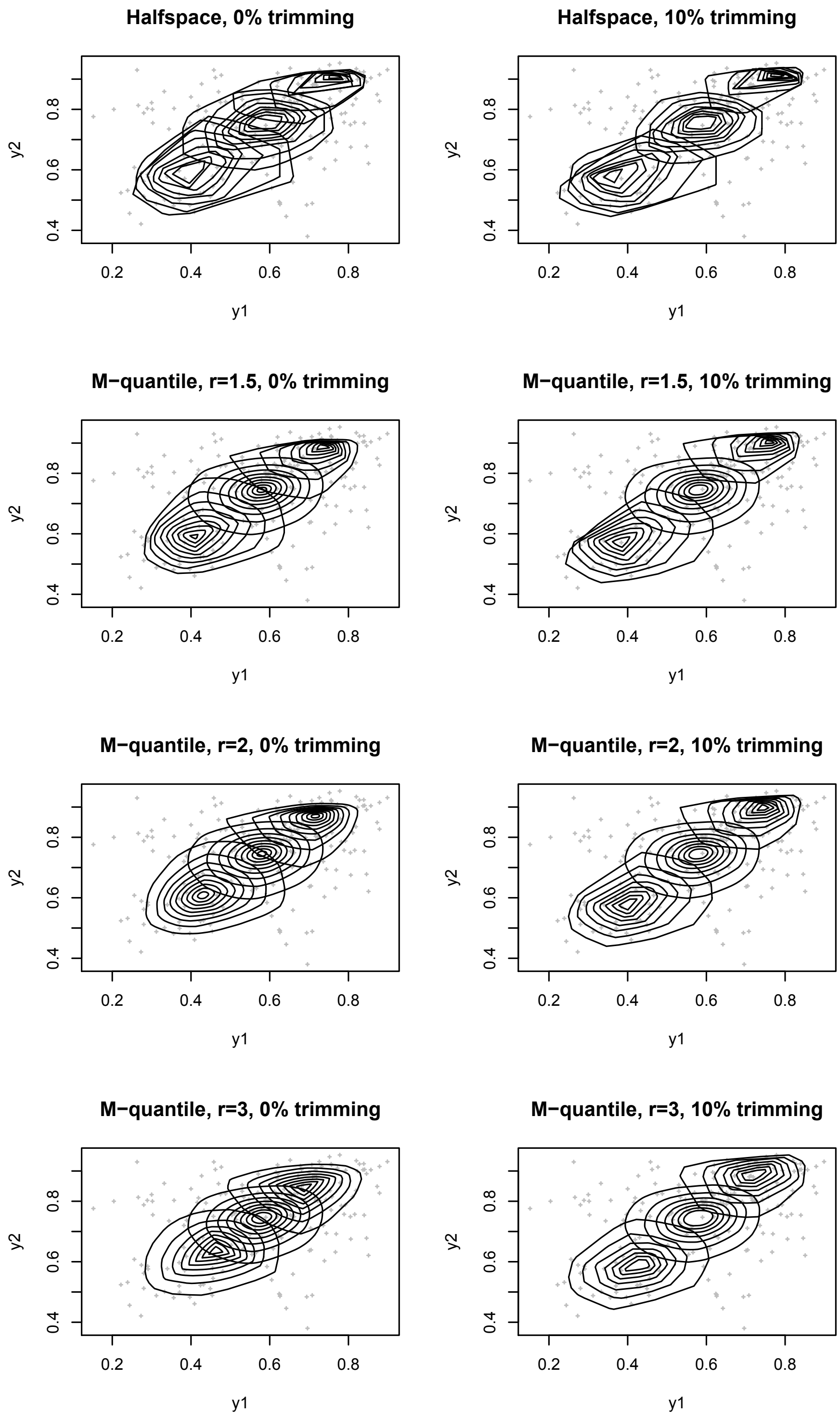

Figure 3: Halfspace and M-quantile conditional regions for various power loss functions.

A Bivariate distorted expectile depth routine

The computation of the exact (undistorted) bivariate expectile depth of a point can be done by means of an algorithm with time complexity $\mathcal{O}(n \log n)$, where n is the sample size, see Cascos and Ochoa (2021). Similarly to the algorithms for the bivariate halfspace and simplicial depths in Rousseeuw and Ruts (1996), such routine moves along the rays with origin in the point whose depth is computed and passing through the points in the dataset, and its time complexity is determined by the one of the algorithm that sorts the angles formed by such rays and some fixed reference ray.

The computation of the distorted expectile depth is slightly more complicated due to the (possibly) different weights awarded to each data point by the distortion function for each univariate projection. Assume that we want to compute the depth of the origin of coordinates $\mathbf{0} \in \mathbb{R}^2$, or alternatively subtract the point whose depth is to be computed from each of the points in the sample to reach that situation. Since all possible orderings of the univariate projections of the sample points must be considered, we use a circular sequence algorithm, which serves quite efficiently for this task.

In first place, the angles formed by each of the $\binom{n}{2}$ straight lines containing each pair of points from the sample with coordinate axis X are computed as well the angles formed by the coordinate axis and the rays from the origin of coordinates passing through the points in the dataset, and they are sorted, which limits the time complexity to $\mathcal{O}(n^2 \log n)$. Then these angles are used to obtain all possible sortings in terms of a univariate projection and the sign of the projection itself. These type of algorithms are commonly used to compute the extreme points of a bivariate central region, see Ruts and Rousseeuw (1996) for the halfspace regions, Dyckerhoff (2000) for the zonoid regions, Cascos (2007) for the expected convex hull regions, or Cascos and Ochoa (2021) for the expectile regions. Nevertheless it has also been used in the computation of some notions of depth, like the bivariate projection depth in Zuo and Lai (2011).

A.1 Preliminaries

Consider a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$, a distortion function g, and $r = 2$. After Theorem 4.3, the sample distorted expectile depth of the origin of coordinates with respect to the previous sample is

$$\mathrm{QD}_n^*(\mathbf{0}) = \left(1 + \sup_{\boldsymbol{u} \in \mathbb{S}^{d-1}} \frac{\sum w_i \langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(i)}, \boldsymbol{u} \rangle_-}{\sum w_i \langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(i)}, \boldsymbol{u} \rangle_+}\right)^{-1}. \tag{12}$$

In the bidimensional setting, $d = 2$, we can write the unit vector $\boldsymbol{u} \in \mathbb{S}^1$ as $\boldsymbol{u} = (\cos\gamma, \sin\gamma)$ for some $\gamma \in [0, 2\pi)$. Consider the two points

$$\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_+} = \sum_{\langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(i)}, \boldsymbol{u}\rangle > 0} w_i \boldsymbol{x}_{\pi_{\boldsymbol{u}}(i)} \quad \text{and} \quad \overline{\boldsymbol{x}}^w_{\boldsymbol{u}_-} = \sum_{\langle \boldsymbol{x}_{\pi_{\boldsymbol{u}}(i)}, \boldsymbol{u}\rangle < 0} -w_i \boldsymbol{x}_{\pi_{\boldsymbol{u}}(i)}\,, \tag{13}$$

which correspond to the weighted sums of the sample points ordered according to their univariate projections times $\boldsymbol{u}$, in first place in the halfspace with inner normal vector $\boldsymbol{u}$ and then in the halfspace with inner normal vector $-\boldsymbol{u}$.

Write now the former points as their norms times a unit vector, which in terms of angles ϕ_+ and ϕ_- correspond to $\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_+} = \|\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_+}\|(\cos\phi_+, \sin\phi_+)$ and $\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_-} = \|\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_-}\|(\cos\phi_-, \sin\phi_-)$, then (12) renders:

$$\mathrm{QD}^*_n(\mathbf{0}) = \left(1 + \sup_{0\le\gamma<2\pi} \frac{\|\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_-}\|}{\|\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_+}\|} \frac{\cos(\gamma - \phi_-)}{\cos(\gamma - \phi_+)}\right)^{-1}. \tag{14}$$

Observe next that $\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_-}$ and $\overline{\boldsymbol{x}}^w_{\boldsymbol{u}_+}$ for $\boldsymbol{u} = (\cos\gamma, \sin\gamma)$ depend on γ, but are constant over each subinterval of $[0, 2\pi)$ at which the sign of each $\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle$ remains unchanged and also the permutation $\pi_{\boldsymbol{u}}$ remains fixed. The second part of the function to be maximized is

$$f(\gamma) = \frac{\cos(\gamma - \phi_-)}{\cos(\gamma - \phi_+)},$$

whose monotonicity depends only on the sign of the sine of $\phi_- - \phi_+$, since the derivative of f is

$$f'(\gamma) = \frac{\sin(\phi_- - \phi_+)}{\cos^2(\gamma - \phi_+)}.$$

This means that the maximum at each subinterval will be attained for γ lying at one of its endpoints.

Finally, the proposed algorithm will consider the partition of $[0, 2\pi)$ with breaking points at the angles for which either $\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle = 0$ for some $i = 1, \dots, n$ or $\langle \boldsymbol{x}_i, \boldsymbol{u}\rangle = \langle \boldsymbol{x}_j, \boldsymbol{u}\rangle$ for some $i \neq j$.

A.2 The routine

The source code for the bivariate distorted expectile depth computation in R is available on the GitHub repository `https://github.com/icascos/expdepth` as function `distexpdepth`.

The input consist in a matrix $\mathbf{X}$ of size $n \times 2$ with all data points in general position (no more than two points in the same straight line) and a distortion function g, while the output is the distorted expectile depth of the origin of coordinates.

Set $w[i] = \widetilde{g}\,(i/n) - \widetilde{g}\,((i-1)/n)$;
Create matrix **ANG** with 3 columns and $n + \binom{n}{2}$ rows;
for $i = 1$ **to** n **do**
  Call **angle** to angle in $(-\pi, \pi]$ between X-axis and vector from origin to $\mathbf{X}[i,]$;
  Set row i of matrix **ANG** as $(i, i, \mathbf{angle})$;
  Set **label**$[i]$ as $+1$ if **angle** in interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and -1 otherwise;
end
for $i = 1$ **to** $n - 1$ **do**
  for $j = i + 1$ **to** n **do**
    Call **angle** to angle in $(-\pi, \pi]$ between X-axis and vector from $\mathbf{X}[i,]$ to $\mathbf{X}[j,]$;
    Set next row of matrix **ANG** as $(i, j, \mathbf{angle})$;
  end
end
Subtract π to angles (matrix **ANG** col. 3) in $(\pi/2, \pi]$ and add π to angles in $(-\pi, -\pi/2)$;
Sort rows of matrix **ANG** with respect to increasing angles (col. 3);
(*) Append $n + \binom{n}{2}$ rows to matrix **ANG** identical to **ANG** except for col. 3 at which π is added to each angle;
Set **Ratio** $= 0$ and as in (13) for $\boldsymbol{u} = (1, 0)$, sort the data points in terms of its first coordinates, $\mathbf{SUM}_+$ is the weighted sum of points in $\mathbf{X}$ labeled as $+1$ with weights $w[i]$ depending on the sorting, and $\mathbf{SUM}_-$ is the corresponding weighted sum for points labeled as -1;
Set $\mathbf{Ratio} = \max\left\{\mathbf{Ratio}, -\frac{\langle \mathbf{SUM}_-, (\cos(\mathbf{ANG}[i,3]+\pi/2), \sin(\mathbf{ANG}[i,3]+\pi/2)\rangle}{\langle \mathbf{SUM}_+, (\cos(\mathbf{ANG}[i,3]+\pi/2), \sin(\mathbf{ANG}[i,3]+\pi/2)\rangle}\right\}$;
for $i = 1$ **to** $2n + 2\binom{n}{2}$ **do**
  if $\mathbf{ANG}[i,1] = \mathbf{ANG}[i,2]$ *(angle defined by one data point)* **then**
    Update $\mathbf{SUM}_+$ adding weighted coordinates of the data point if labeled as -1 and subtracting weighted coordinates of the data point if labeled as $+1$;
    Update $\mathbf{SUM}_-$ adding weighted coordinates of the data point if labeled as $+1$ and subtracting weighted coordinates of the data point if labeled as -1;
  end
  if $\mathbf{ANG}[i,1] \neq \mathbf{ANG}[i,2]$ *(angle defined by two data points)* **then**
    if *point* $\mathbf{X}[i,]$ *labeled as* $+1$ **then**
      Update $\mathbf{SUM}_+$ interchanging the weights of points $\mathbf{ANG}[i,1]$ and $\mathbf{ANG}[i,2]$;
    end
    if *point* $\mathbf{X}[i,]$ *labeled as* -1 **then**
      Update $\mathbf{SUM}_+$ interchanging the weights of points $\mathbf{ANG}[i,1]$ and $\mathbf{ANG}[i,2]$;
    end
  end
  Set $\mathbf{Ratio} = \max\left\{\mathbf{Ratio}, -\frac{\langle \mathbf{SUM}_-, (\cos(\mathbf{ANG}[i,3]+\pi/2), \sin(\mathbf{ANG}[i,3]+\pi/2)\rangle}{\langle \mathbf{SUM}_+, (\cos(\mathbf{ANG}[i,3]+\pi/2), \sin(\mathbf{ANG}[i,3]+\pi/2)\rangle}\right\}$;
end
Return the distorted expectile depth as: $(1 + \mathbf{Ratio})^{-1}$, see (14).

Algorithm 1: Bivariate distorted expectile depth

Remark A.1. Since there are n points, $n + \binom{n}{2}$ angles must be considered, hence computing and ordering the angles of each pair of points can be done with an algorithm of time complexity $\mathcal{O}(n^2 \log n)$. The main loop is run precisely $2n + 2\binom{n}{2}$ times, but the number of operations at each iteration does not depend on the sample size n, since only one point is being added, subtracted or reweighted.

Remark A.2. If the distortion function is symmetric, $g = \widetilde{g}$, it is possible to skip step (*) as long as every time we update **Ratio**, we also compare with the inverse of the quotient of the projections of the weighted sums of the data points at each of the two halfplanes.

References

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

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

I. Cascos. The expected convex hull trimmed regions of a sample. Comput. Statist. **22**, 2007, 557–569.

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

I. Cascos and M. Ochoa (2021) Expectile depth: theory and computation for bivariate datasets. Journal of Multivariate Analysis, **184**, 2021.

A. Daouia and D. Paindaveine. (2019). From halfspace M-depth to multiple-output expectile regression. `https://arxiv.org/abs/1905.12718`.

D. Denneberg. (1994). Non-additive measure and integral. Kluwer Academic Publishers.

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

M. Hallin, D. Paindaveine, and M. Siman. Multivariate quantiles and multiple output regression quantiles: from L_1 optimization to halfspace depth. The Annals of Statistics, **Vol. 38, No.2**, 2010 635–669.

M. Hallin, Z. Lu, D. Paindaveine, and M. Siman. Local bilinear multiple-output quantile/depth regression. Bernoulli, **21(3)**, 2015, 1435–1466.

R. Koenker. Quantile regression. Cambridge University Press, 2005.

R. Koenker and G. Basset. Regression Quantiles. Econometrica, **46**, 1978, 33–50.

G. Koshevoy and K. Mosler. Zonoid trimming for multivariate distributions. Ann. Statist., 1997, **25**, 1998–2017.

R.Y. Liu. On a notion of data depth based on random simplices. Ann. Statist., **18**, 1990, 405–414.

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

R.Y. Liu and K. Singh. A Quality Index Based on Data Depth and Multivariate Rank Tests. J. Am. Statist. Assoc. **88**, 1993, 252–260.

L. Merlo, L. Petrella, N. Salvati, and N. Tzavidis. Marginal M-quantile regression for multivariate dependent data. Comput. Statist. Data Anal. **173**, 2022, 107500.

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

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

O. Pokotylo, P. Mozharovskyi, R. Dyckerhoff, and S. Nagy, (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`.

P. Rousseeuw and I. Ruts. Algorithm AS 307: Bivariate Location Depth. J. Roy. Statist. Soc. Ser. C **45**, 1996, 516–526.

P. Rousseeuw and I. Ruts. The depth function of a population distribution. Metrika, **49**, 1999, 213–244.

I. Ruts and P. Rousseeuw. Computing depth contours of bivariate point clouds. Comput. Statist. Data. Anal. **23**, 1996, 153–168.

J. Tukey. Mathematics and the picturing of data. Proc. Int. Congress of Math. (Vancouver, 1974) Vol. 2, 1975, 523–531.

Y. Zuo and S. Lai. Exact computation of bivariate projection depth and the Stahel–Donoho estimator. Comput. Statist. Data Anal. **55**, 2011, 1173–1179

Y. Zuo and R. Serfling. General notions of statistical depth function. Ann. Statist. **28**, 2000, 461–482.

Y. Zuo and R. Serfling. Structural properties and convergence results for contours of sample statistical depth functions. Ann. Statist. **28**, 2000, 483–499.

Y. Zuo and R. Serfling. On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. J. Statist. Plann. Inference, **84**, 2000, 55–79.