Expectile Based Ranking for Multi-Criteria Decision Making

Ha Thi Khanh Linh[a] **and Andreas H Hamel**[b;*]

[a]Free University of Bozen, Faculty for Economics and Management
[b]Free University of Bozen, Faculty for Economics and Management

Abstract. Based on a set-valued generalization of expectiles, known from regression analysis and statistics, new categorization and ranking procedures for multi-dimensional data points are proposed. The ranking is extended to sets of alternatives.

1 Introduction

The ranking problem in Multi-Criteria Decision Analysis is to find a complete order for a number of multi-dimensional alternatives, i.e., order them in such a way that one can always say which of two given alternatives is ranked higher or at least equal than the other. In most cases, there is an underlying non-complete order relation which is very intiutive: the problem is that for this relation some pairs of alternatives are not comparable. The most popular example is the component-wise order which is used when an alternative is described via several numerical criteria. In this case, one alternative can be better with respect to some of the criteria than another, but worse with respect to other criteria.

The ranking, i.e., the transition from a non-complete to a complete order, usually comes with a loss of information: one cannot tell anymore if a higher ranked alternative was already better with respect to the original order or if the higher rank is due to the ranking procedure itself.

Similarly, if one wants to categorize multi-dimensional data points in, say, the "good" ones and the "bad" ones, for example, with respect to the component-wise order in $d > 1$ dimensions, it might happen that some points are neither "good," nor "bad," but just not categorizable (yet). Such a categorization is certainly a valid goal and useful for learning procedures when an intuitive order relation for the multi-dimensional data points is present as for the evaluations of objects like songs, movies, projects etc. with respect to several criteria.

The aim of this paper is threefold. First, a categorization of multi-dimensional data points is proposed which is based on a set-valued generalization of expectiles, a statistical concepts which can be understood as be in between the expected value and a quantile—both used to order one-dimensional data in statistics. Secondly, a new ranking procedure for multi-dimensional alternatives is introduced which is based on set-valued expectiles. Finally, the ranking procedure is extended to rankings of sets of alternatives in the spirit of [6, 7].

An open problem is formulated about the coupling of the suggested procedure with a dimension reduction technique to avoid the 'curse of dimension' and remove correlations. Results in this direction would open the path for using the proposed method effectively in Machine Learning and Data Analysis.

* Corresponding Author. Email: andreas.hamel@unibz.it

2 Notation and basic problem

The symbols $\mathbb{N}$ and $\mathbb{R}$ are used for the sets of natural (including 0) and real numbers, respectively.

Let $d \in \mathbb{N}\backslash\{0,1\}$ and $C \subseteq \mathbb{R}^d$ be a proper closed convex cone. This means that C is a closed set satisfying $sC = C$ for all $s \geq 0$, $C + C = C$, $C \not\in \{\varnothing, \mathbb{R}^d\}$. It generates a vector preorder $\leq_C$ (a reflexive and transitive relation which is compatible with addition and multiplication with non-negative numbers) via $y \leq_C z$ iff $z - y \in C$. Special cases are the zero cone $C = \{0\}$ and closed half-spaces $C = H^+(w) := \{z \in \mathbb{R}^d \mid w^T z \geq 0\}$ for $w \in \mathbb{R}^d\backslash\{0\}$ as well as, of course, $C = \mathbb{R}^d_+$ which generates the component-wise order for d-dimensional vectors. The symbol $y <_C z$ is used for $z - y \in \operatorname{int} C$ where $\operatorname{int} C \neq \varnothing$ is assumed.

Let $N \in \mathbb{N}\backslash\{0,1\}$ alternatives be given, i.e., a set $X = \{x^1, \dots, x^N\} \subseteq \mathbb{R}^d$. The problems addressed in this paper are

- clustering/categorizing the points in X into points which dominate many other points, points which are dominated by many others, both with respect to the order $\leq_C$, and those to which neither one criterion applies where it has to be specified what is understood by "many,"
- rank the points in X by means of a function $r\colon X \to \mathbb{R}$ which satisfies
$$y \leq_C z \quad \Rightarrow \quad r(y) \leq r(z), \tag{1}$$
- rank sets $A, B \subseteq X$ of alternatives instead of single points.

3 Categorization via expectiles

Univariate expectiles are defined in [5] as unique solutions of minimization problems for asymmetric quadratic expected loss functions; such minimization problems facilitate the so-called expectile regression. The necessary optimality condition for this problem gives the α-expectile $e_\alpha(X)$ of X with $0 < \alpha < 1$ as unique solution of the equation
$$\alpha \mathbb{E}(X - e)_+ = (1 - \alpha)\mathbb{E}(X - e)_-, \tag{NOC}$$
see [1, 5]. This equation makes sense for any random variable $X\colon \Omega \to \mathbb{R}$ for which the expected value $\mathbb{E}(X)$ exists. The positive and the negative part of a number $x \in \mathbb{R}$ used in (NOC) are defined as $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$.

The sample version of the above equation with the data points $x_1, \dots, x_N \in \mathbb{R}$ becomes

$$\alpha \sum_{x_n > e} (x_n - e) = (1-\alpha) \sum_{x_n < e} (e - x_n). \tag{2}$$

It can been understood as finding a number e which makes the sum of positive differences between data points and e weighted with α equal to the negative sum of the negative differences weighted with $1-\alpha$. One may see from this that this number e is just the average of the x-values if $\alpha = \frac{1}{2}$ and that a greater α produces a greater e.

The uniques solution of (2) is called sample expectile of the data set $X = \{x_1, \dots, x_N\}$ and is denoted by $e_{N,\alpha}(X)$. The sample expectile divides the data points into two categories, the ones which are less than $e_{N,\alpha}(X)$ and the ones which are greater or equal.

If the data points are multi-dimensional, i.e., $X = \{x^1, \dots, x^N\} \subseteq \mathbb{R}^d$ with $d \geq 2$, it is much harder to define the sample expectile which incorporates the order $\leq_C$ generated by C such that the data points can be categorized. A recent proposal is from [2] and gives a set-valued analog to $e_{N,\alpha}(X)$. The first step is to define the expectile for projected data points. For $w \in C^+ \backslash \{0\}$ the unique solution $e = e_{N,\alpha,w}(X)$ of the equation

$$\alpha \sum_{w^\top x^n > e} (w^\top x^n - e) = (1-\alpha) \sum_{w^\top x^n < e} (e - w^\top x^n) \tag{3}$$

is the α-expectile for the set of numbers $w^\top X := \{w^\top x^1, \dots, w^\top x^N\}$ where $w^\top x$ denotes the usual scalar product of the two vectors $w, x \in \mathbb{R}^d$. The following definition is a sample version of the one from [2].

Definition 1 *For $0 < \alpha < 1$, the set*

$$E^{N,\alpha}_{-C}(X) = \bigcap_{w \in C+} \{z \in \mathbb{R}^d \mid w^\top z \leq e_{N,\alpha,w}(X)\} \tag{4}$$

is called the downward cone α-expectile set and the set

$$E^{N,1-\alpha}_{C}(X) = \bigcap_{w \in C+} \{z \in \mathbb{R}^d \mid w^\top z \geq e_{N,1-\alpha,w}(X)\} \tag{5}$$

the upward cone α-expectile set of X.

Indeed, the sets $E^{N,\alpha}_{-C}(X)$, $E^{N,1-\alpha}_{C}(X)$ can be used as tools for a categorization of the data points in X. However, as the following example shows a third category of data points appears: in addition to the "good" points and the "bad" points, there are points which are neither good, nor bad and cannot yet be categorized. This is of course due to the fact that the order relation $\leq_C$ is not complete in general which means that there are non-comparable points—in contrast to the usual $\leq$-order for the real numbers. Figure 1 shows an example with data from a currency exchange market with transaction costs between the three currencies. This means that the buyer's and the seller's price for an exchange, say, EUR into USD, are different which leads to a convex cone, called the solvency cone in finance. The red points in the figure stand for holdings of the UK central bank's asset reserves in three currencies: US Dollar (USD), Euro (EUR) and British sterling (GBP) in millions. The data are taken from the website of Bank of England (BOE), for 39 periods (quarters) from September 2013 to December 2022 and can be made available upon request to the authors.

For $\alpha = 0.25$, the upper right part in Figure 1, taken from [2], is the set $E^{N,1-\alpha}_{C}(X)$ while the lower left is the set $E^{N,\alpha}_{-C}(X)$ (only the corresponding boundaries are shown) each containing some of the data points. Between these two sets, the "non-categorizable" points can be found. An increase of α would move these two sets further apart, thus making more points "non-categorizable." A decrease, likewise, would move the sets closer together. The determination of α thus is a subjective decision and depends on the purpose of the categorization process: if one wants to find "really good" and "really bad" points, α would be chosen closer to 1 in which case one does not care that many points cannot be categorized. If, on the other hand, one wants to categorize as many points as possible, one would choose α closer to $\frac{1}{2}$.

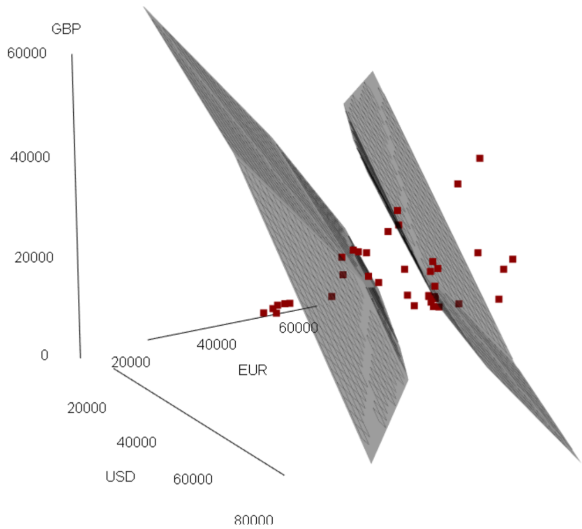

Figure 1. 3D categorization via expectile sets for financial positions.

A few basic properties of the categorizing sets are collected in the following result.

Proposition 2 *The expectile category sets satisfy*

(i) $E^{N,\alpha}_{-C}(X) - C = E^{N,\alpha}_{-C}(X)$, $E^{N,1-\alpha}_{C}(X) + C = E^{N,1-\alpha}_{C}(X)$,

(ii) $X \leq_C Y$ implies $E^{N,\alpha}_{-C}(X) \subseteq E^{N,\alpha}_{-C}(Y)$, $E^{N,1-\alpha}_{C}(X) \supseteq E^{N,1-\alpha}_{C}(Y)$.

(iii) $E^{N,\alpha}_{-C}(AX + b) = AE^{N,\alpha}_{-C}(X) + b$, $E^{N,1-\alpha}_{C}(AX + b) = E^{N,1-\alpha}_{C}(X) + b$ for all $b \in \mathbb{R}^d$ and invertible $d \times d$-matrices A.

(iv) if $0 < \alpha \leq \beta < 1$, then $E^{N,\alpha}_{-C}(X) \subseteq E^{N,\beta}_{-C}(X)$, $E^{N,1-\alpha}_{C}(X) \subseteq E^{N,1-\beta}_{C}(X)$.

The proof of this result along with more properties of expectile sets can be found in [2].

4 Expectile rank functions for multi-criteria decision making

The following definitions provides two expectile-based functions which turn out to be useful for ranking procedures.

Definition 3 *The functions $D_{-C}(\cdot; X), D_C(\cdot; X) \colon \mathbb{R}^d \to [0,1]$ defined by*

$$D_{-C}(z; X) = \inf\{\alpha \in (0,1) \mid z \in E^{N,\alpha}_{-C}(X)\}$$
$$D_{C}(z; X) = \sup\{\beta \in (0,1) \mid z \in E^{N,\beta}_{C}(X)\}$$

are called downward and upward expectile rank function generated by X, respectively.

Proposition 2 (iv) ensures that the set $E_{-C}^{N;\alpha}(X)$ shrinks if α decreases. Thus, $D_{-C}(z;X)$ yields the least value of α such that z is still caught by $E_{-C}^{N,\alpha}(X)$, and one can expect that z can be found on the boundary of $E_{-C}^{N,\alpha}(X)$ for $\alpha = D_{-C}(z;X)$. This is illustrated in Figure 2: the dashed lines are the boundaries of the sets $E_{-C}^{N,\alpha}$ for different α's and the level sets of $D_{-C}(\cdot;X)$ at the same time. A similar interpretation applies to $D_C(z;X)$ with reverse direction of movement: $E_C^{N,\beta}(X)$ shrinks if β increases.

The expectile rank functions indeed satisfy (1), i.e., they can be used to rank the points in X.

Proposition 4 *If $y, z \in \mathbb{R}^d$ satisfy $y \leq_C z$, then $D_{-C}(y;X) \leq D_{-C}(z;X)$ and $D_C(y;X) \leq D_C(z;X)$.*

PROOF. The proof is shown for D_C, the one for D_{-C} runs parallel. Assume $y \leq_C z$, i.e., $z - y \in C$. Proposition 2 (i) implies

$$\left\{\beta \in (0,1) \mid y \in E_C^{N,\beta}(X)\right\} \subseteq \left\{\beta \in (0,1) \mid z \in E_C^{N,\beta}(X)\right\}.$$

since $z = y + (z-y) \in y + C$. Taking the supremum on both sides of this inclusion yields the result.

Figure 2 shows how to spot the downward and upward expectile ranks of a point in $\mathbb{R}^2$ for a two dimesional data set and $C = \mathbb{R}^2_+$. The data points in X are generated as a sample of a bivariate distribution taken via the Gumbel copula in which the two marginal distributions are the normal distribution $N(7,4)$ and the gamma distribution $\Gamma(\alpha = 4, \beta = 3)$. The downward expectile rank of the "red" point z is $D_{-\mathbb{R}^2_+}(z;X) = 0.35$ and its upward expectile rank is $D_{\mathbb{R}^2_+}(z;X) = 0.2$. This example is taken from [2]. One may also realize from this picture that data points which are not comparable component-wise can have the same (such as z and y) or (very) different values of the ranking functions $D_{-\mathbb{R}^2_+}$, $D_{\mathbb{R}^2_+}$.

5 Ranking of sets of alternatives

Let $A, B \subseteq \mathbb{R}^d$ be two finite sets of alternatives, for example, the outcomes of two runs of algorithms for a multi-objective optimization problem. If maximization is the goal, then A is certainly preferred over B, written $A \succeq_C B$, if

$$\forall b \in B,\ \exists a \in A\colon\ b \leq_C a.$$

The relation $\succeq_C$ on the power set $\mathcal{P}(\mathbb{R}^d)$ is reflexive, transitive, but not antisymmetric even if $\leq_C$ is. See [3] for this fact and many more details and references. Moreover, it is of course not complete which means that it is the rule rather than the exception that two sets $A, B \subseteq \mathbb{R}^d$ neither satisfy $A \succeq_C B$, nor $B \succeq_C A$. This relation has been used in [6] to compare outcomes of evolutionary algorithms for multi-objective optimization problems.

A ranking which maintains $\succeq_C$ is therefore desirable, i.e., a function $R\colon \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ which satsifies

$$A \succeq_C B \quad \Rightarrow \quad R(A) \geq R(B). \tag{6}$$

Such a ranking is given in the next definition.

Definition 5 *The functions $D_C^{\vartriangle}(\cdot;X), D_C^{\triangledown}(\cdot;X)\colon \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ defined by*

$$D_C^{\vartriangle}(A;X) = \sup_{a \in A} D_C(a;X) \text{ and } D_C^{\triangledown}(A;X) = \inf_{a \in A} D_C(a;X)$$

is called the upper and lower expectile set rank function, respectively.

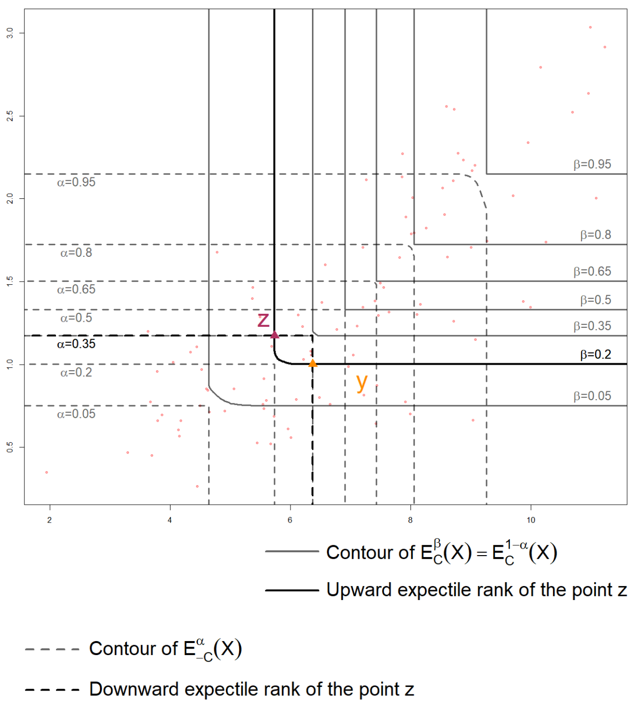

Figure 2. Downward and upward expectile ranks of points in 2D.

Proposition 6 *The function $D_C^{\vartriangle}(\cdot;X)$ satisfies (6), i.e., it is monotone with respect to $\succeq_C$.*

PROOF. Since the upward expectile rank function D_C is monotone (see Proposition 4), the definition of $\succeq_C$ gives

$$\forall b \in B,\ \exists a \in A\colon\ D_C(b;X) \leq D_C(a;X).$$

Taking the supremum over B and A on the left and right side of this inequality, respectively, yields the desired inequality.

Remark 7 *If minimization is the goal, the relation $\succeq_C$ is not appropriate anymore. Instead, one can use $A \succeq^C B$ defined by*

$$\forall a \in A,\ \exists b \in B\colon\ b \leq_C a.$$

It is well-known that $\succeq^C$ also is reflexive and transitive, but different from $\succeq_C$, see [3]. The reader may easily verify that the lower expectile rank function $D_C^{\triangledown}(\cdot;X)$ is monotone w.r.t. $\succeq^C$. The relation $\succeq^C$ has also been used for ranking the outcome sets of evolutionary algorithms for multi-objective optimization problems [7].

6 Conclusion and open problems

A new procedure for categorization and ranking of multi-dimensional data points based on expectiles is proposed. This is another instance of transferring a statistical functions into a ranking function which complements the idea of [4] where quantile functions have been used in a similar way.

However, in many applications, the dimension d is large and there is correlation present among the data points. For example, data points can be essentially confined to lower dimensional subspaces of $\mathbb{R}^d$.

Moreover, a visualization of data features is only possible in dimensions 2 or 3. Therefore, it is highly desirable to couple a dimension reduction technique such as principal component or factor analysis to reduce the dimension to a number $1 \leq m < d$ and adjust the expectile sets as well as the expectile rank functions for this situation. This is an open research problem which will be pursued by the authors in the future.

References

[1] Fabio Bellini, Bernhard Klar, Alfred Müller, and Emanuela Rosazza Gianin, 'Generalized quantiles as risk measures', *Insurance: Mathematics and Economics*, **54**, 41–48, (2014).

[2] Thi Khanh Linh Ha and Andreas H Hamel, 'Set-valued expectiles for ordered data analysis', *arXiv preprint arXiv:2312.09930*, (2023).

[3] Andreas H Hamel, Frank Heyde, Andreas Löhne, Birgit Rudloff, and Carola Schrage, 'Set optimization?a rather short introduction', in *Set Optimization and Applications-The State of the Art: From Set Relations to Set-Valued Risk Measures*, pp. 65–141. Springer, (2015).

[4] Andreas H Hamel and Daniel Kostner, 'Cone ranking for multi-criteria decision making', *arXiv preprint arXiv:2312.03006*, (2023).

[5] Whitney K Newey and James L. Powell, 'Asymmetric least squares estimation and testing', *Econometrica*, **55**, 819–847, (1987).

[6] Eckart Zitzler and Lothar Thiele, 'Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach', *IEEE transactions on Evolutionary Computation*, **3**(4), 257–271, (1999).

[7] Eckart Zitzler, Lothar Thiele, and Johannes Bader, 'On set-based multiobjective optimization', *IEEE Transactions on Evolutionary Computation*, **14**(1), 58–79, (2009).