

Stochastically ordered aggregation operators

*Original*

Stochastically ordered aggregation operators / Juan Baz, Juan; Pellerey, Franco; Díaz, Irene; Montes, Susana a. - In: INTERNATIONAL JOURNAL OF APPROXIMATE REASONING. - ISSN 1873-4731. - ELETTRONICO. - 166:(2024), pp. 1-17. [10.1016/j.ijar.2023.109110]

*Availability:*

This version is available at: 11583/2985285 since: 2024-02-25T14:51:37Z

*Publisher:*

Elsevier

*Published*

DOI:10.1016/j.ijar.2023.109110

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



# Stochastically ordered aggregation operators

Juan Baz<sup>a,\*</sup>, Franco Pellerey<sup>b</sup>, Irene Díaz<sup>c</sup>, Susana Montes<sup>a</sup>

<sup>a</sup> Department of Statistics and I.O. and Didactic of Mathematics, University of Oviedo, Calle Federico Garcia Lorca 18, Oviedo, 33007, Asturias, Spain

<sup>b</sup> Department of Mathematical Sciences "G.L. Lagrange", Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Piedmont, Italy

<sup>c</sup> Department of Computer Science, University of Oviedo, Campus de Viesques, Gijón, 33007, Asturias, Spain

## ARTICLE INFO

### Keywords:

Aggregation operators  
Stochastic orders  
Aggregation of random variables  
Copulas

## ABSTRACT

In aggregation theory, there exists a large number of aggregation functions that are defined in terms of rearrangements in increasing order of the arguments. Prominent examples are the Ordered Weighted Operator and the Choquet and Sugeno integrals. Following a probability approach, ordering random variables by means of stochastic orders can be also a way to define aggregations of random variables. However, stochastic orders are not total orders, thus pairs of incomparable distributions can appear. This paper is focused on the definition of aggregations of random variables that take into account the stochastic ordination of the components of the input random vectors. Three alternatives are presented, the first one by using expected values and admissible permutations, then a modification for multivariate Gaussian random vectors and a third one that involves a transformation of the initial random vectors in new ones whose components are ordered with respect to the usual stochastic order. A deep theoretical study of the properties of all the proposals is made. A practical example regarding temperature prediction is provided

## 1. Introduction

Aggregation theory is devoted to the fusion of information of different sources in an unique element that summarizes all the available information. Working with real numbers, an aggregation function takes a tuple of dimension  $n$  of a real interval and returns another value in the same interval, fulfilling monotonicity and some boundary conditions over the ends of the interval [12].

Aggregation functions are widely used in the context of Data Analysis [3,16,23]. The most classical approach to model data is to see them as realizations of a collection of random variables, as considered in Statistics [24]. Then, it is reasonable to combine Aggregation Theory and Probability Theory to study the aggregation of data. In this direction, the concept of aggregations of random variables was defined in [1], as functions that take random vectors and return random variables. The monotonicity and boundary conditions are redefined by using stochastic orders.

In classical Aggregation Theory, ordering the values of the input vector is a quite common way to define aggregation functions. Prominent examples are the OWA operator [30], the IOWA operator [31] and the Choquet and Sugeno integrals [12]. Moving to the stochastic setting, to generalize such operators it is reasonable to define an aggregation of random variables that takes into account the order of the input random variables. However, as pointed out by Yager in [32], the usual stochastic order (whose definition is

\* Corresponding author.

E-mail address: [bazjuan@uniovi.es](mailto:bazjuan@uniovi.es) (J. Baz).

recalled below) is not a total order in the set of probability distributions (see [25]), thus pairs of incomparable random variables may appear.

We devote this paper to the definition of aggregation of random variables based on the stochastic ordination of the components of the input random vector. With this approach, the whole distribution of the random variables can be used to order them, while in classical aggregation theory only the observations of realizations are used for such ordination.

The definition of aggregation functions based on partial orders has been studied in detail in the literature in the case of aggregation of vectors of real numbers [22]. Here we consider the same problem but dealing with random variables, approaching at the incomparability problem by two different techniques. Firstly, we order the random variables in terms of their means and considering admissible permutations (see [21]) in the case of ties. A slight modification of this approach has good properties when working with multivariate Gaussian random vectors. The second technique transforms the initial random vector in a new one in which its components are ordered in the usual stochastic order, following a similar approach as the one proposed by Yager in [32].

The remainder of the paper is organized as follows. The basic concepts about stochastic orders, copulas, aggregation of random variables and ordination of arguments in Aggregation Theory are presented in Section 2. Section 3 is devoted to the definition and the study of the Mean Ordered Weighted Averaging Operator and its modification for multivariate Gaussian random vectors. The Stochastically Ordered Aggregation Operator is presented in Section 4, as well as several results characterizing its properties. A practical example that illustrates its benefits when used in a problem of prediction is provided in Section 5, and the conclusions are discussed in Section 6.

## 2. Preliminaries

We devote this section to introducing the basic concepts needed for the development of the paper. In particular, we will focus on stochastic orders, copulas, aggregation of random variables and, finally, we will discuss some examples of ordered aggregations in the literature.

### 2.1. Stochastic orders and copulas

Comparing probability distributions is a common problem in several areas such as Reliability Theory, Queuing Theory or Statistics [25]. A huge amount of relations between distributions, known as stochastic orders, have been defined in this regard. These orders can be extended to random vectors straightforwardly, since any random vector induces a probability distribution. Different stochastic orders will be considered in this paper, mainly defined in terms of expectations of particular type of functions. In the following, we assume that all random vectors are of dimension  $n$ . In addition, we will use increasing in a non strict sense.

**Definition 1.** [25] Let  $\vec{X}$  and  $\vec{Y}$  be two random vectors. Then, if  $E[\varphi(\vec{X})] \leq E[\varphi(\vec{Y})]$  for any measurable increasing [convex, increasing convex] function  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $E[\varphi(\vec{X})]$  and  $E[\varphi(\vec{Y})]$  exist,  $\vec{X}$  is said to be smaller than or equal to  $\vec{Y}$  in the usual [convex, increasing convex] stochastic order and it is denoted as  $\vec{X} \leq_{st} [\leq_{cx}, \leq_{icx}] \vec{Y}$ .

For the cases in which  $\vec{X} \leq_{st} [\leq_{cx}, \leq_{icx}] \vec{Y}$  and  $\vec{Y} \leq_{st} [\leq_{cx}, \leq_{icx}] \vec{X}$  are both true, we will use the notation  $\vec{X} =_{st} [=_{cx}, =_{icx}] \vec{Y}$ . In all these cases, equality in the stochastic orders is equivalent to *equality in distribution*, that is,  $P(\vec{X} \in A) = P(\vec{Y} \in A)$  for any measurable set  $A \subseteq \mathbb{R}^n$ .

The usual stochastic order has the following characterization in terms of random vectors defined in the same probability space, which will be useful later. Here and in the rest of the paper the inequality between vectors means that all components of the two vectors  $\vec{x}$  and  $\vec{y}$  are ordered accordingly.

**Theorem 2.** [25] Let  $\vec{X}$  and  $\vec{Y}$  be two random vectors.  $\vec{X} \leq_{st} \vec{Y}$  if and only if there exist two random vectors  $\hat{\vec{X}}$  and  $\hat{\vec{Y}}$  defined in the same probability space such that  $\hat{\vec{X}} =_{st} \vec{X}$ ,  $\hat{\vec{Y}} =_{st} \vec{Y}$  and  $P(\hat{\vec{X}} \leq \hat{\vec{Y}}) = 1$ .

The property  $P(\vec{X} \leq \vec{Y}) = 1$  is known as to be *almost surely smaller* than or equal and will be denoted as  $\vec{X} \leq_{a.s.} \vec{Y}$ . As a consequence of the latter result, whenever  $\vec{X}$  and  $\vec{Y}$  are defined in the same probability space,  $\vec{X} \leq_{a.s.} \vec{Y} \implies \vec{X} \leq_{st} \vec{Y}$ .

In the univariate case, the usual stochastic order is equivalent to the pointwise ordination of the cumulative distribution functions, that is,  $X \leq_{st} Y \iff F_X(t) \geq F_Y(t) \forall t \in \mathbb{R}$  (see [25]).

We want to recall that  $\vec{X} \leq_{st} [\leq_{cx}, \leq_{icx}] \vec{Y}$  implies  $X_i \leq_{st} [\leq_{cx}, \leq_{icx}] Y_i \forall i \in \{1, \dots, n\}$  but in general the other implication does not hold [25]. The equivalence holds in special cases, see Proposition 7.

In the case of multivariate Gaussian distributions, the latter stochastic orders can be easily characterized. These characterizations will be of interest in Subsection 3.2.

**Theorem 3.** [25] Let  $\vec{X}$  and  $\vec{Y}$  be two random vectors having multivariate Gaussian distributions with, respectively, mean vectors  $\vec{\mu}_X$  and  $\vec{\mu}_Y$  and covariance matrices  $\Sigma_X$  and  $\Sigma_Y$ . Then:

- $\vec{X} \leq_{st} \vec{Y}$  if and only if  $\vec{\mu}_X \leq \vec{\mu}_Y$  and  $\Sigma_X = \Sigma_Y$ .

- $\vec{X} \leq_{cx} \vec{Y}$  if and only if  $\vec{\mu}_X = \vec{\mu}_Y$  and  $\Sigma_Y - \Sigma_X$  is a positive semi-definite matrix.
- $\vec{X} \leq_{icx} \vec{Y}$  if and only if  $\vec{\mu}_X \leq \vec{\mu}_Y$  and  $\Sigma_Y - \Sigma_X$  is a positive semi-definite matrix.

We want to recall that a matrix  $M$  of dimension  $n \times n$  is *positive semi-definite* if  $\vec{v}^T M \vec{v} \geq 0$  for any vector  $\vec{v} \in \mathbb{R}^n$ . We will denote the set of positive semi-definite matrices as  $\mathcal{P}$ .

**Definition 4.** [29] Let  $M_1$  and  $M_2$  be two positive semi-definite matrices. Then, is said that  $M_1$  is smaller than or equal to  $M_2$  in the **Loewner order** if  $M_2 - M_1$  is positive semi-definite. It will be denoted as  $M_1 \leq_{Lo} M_2$ .

In addition to these orders, the case of just considering the mean vectors, i.e.  $E[\vec{X}] \leq E[\vec{Y}]$  (component-wise), will be used in Section 3. It will be denoted as  $\vec{X} \leq_E \vec{Y}$ .

The concept of copulas will be also very useful in the development of the paper. The distribution of any random vector can be separated in two parts, the marginal distributions and the copula. The marginals are the distributions of the components of the random vector and the copula describes the information about the dependence between them. The class of copula functions is defined as follows:

**Definition 5.** [20] A function  $C : [0, 1]^n \rightarrow [0, 1]$  is an  **$n$ -copula** if the following conditions are fulfilled:

- For every  $\vec{x} \in [0, 1]^n$ ,  $C(\vec{x}) = 0$  if there exists  $i \in \{1, \dots, n\}$  such that  $x_i = 0$ .
- For every  $\vec{x} \in [0, 1]^n$ ,  $C(\vec{x}) = x_i$  if  $x_j = 1$  for every  $j \neq i$ .
- For every  $\vec{x}, \vec{y} \in [0, 1]^n$  such that  $\vec{x} \leq \vec{y}$  it holds,

$$\sum_{\vec{v} \in P_{[\vec{x}, \vec{y}]}} \text{sgn}(\vec{v})C(\vec{v}) \geq 0,$$

where  $[\vec{x}, \vec{y}]$  is the  $n$ -dimensional prism with corner vertices  $\vec{x}$  and  $\vec{y}$ ,  $P_{[\vec{x}, \vec{y}]}$  denotes the set of vertices of  $[\vec{x}, \vec{y}]$  and

$$\text{sgn}(\vec{v}) = \begin{cases} 1 & \text{if } v_k = x_k \text{ for an even number of } k\text{'s} \\ -1 & \text{if } v_k = x_k \text{ for an odd number of } k\text{'s.} \end{cases}$$

Specifically, it is always possible to represent any multivariate distribution function as a composition of a copula and the distribution functions of the components of the random vectors. This outcome is formally recognized as Sklar’s Theorem, and it will be recalled in Theorem 3. Throughout this work, for the sake of brevity, we will use the term copula in place of  $n$ -copula when the dimension  $n$  is clear.

**Theorem 6.** [26] Let  $\vec{X}$  be a random vector with marginals distributions  $F_1, \dots, F_n$  and joint distribution function  $F$ . Then, it exists a copula  $C : [0, 1]^n \rightarrow [0, 1]$  such that

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

Prominent examples of copulas are the product copula and the minimum copula which are associated, respectively, with independent and comonotone (perfectly positive dependent) random variables. Two random vectors are said to *have the same copula* if the copula of Theorem 6 can be chosen to be the same.

It must be recalled that whenever the marginal distributions  $F_1, \dots, F_n$  are continuous then the copula  $C$  is unique, while in the other cases there exists a unique subcopula  $C'$  defined on  $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$  such that the equation in Theorem 6 holds, where  $\text{Ran}F_i$  denotes the range of  $F_i$  (see, for instance, [8,11]).

It is well known and easy to verify that in the first case the vector  $(U_1, \dots, U_n) = (F_1(X_1), \dots, F_n(X_n))$  has the unique copula and its joint distribution has uniform marginals, while in the second case the distribution of the vector  $(U_1, \dots, U_n)$ , which assume values in  $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$ , is entirely defined by the subcopula  $C'$ .

**Example 1.** Let  $\vec{X} = (X_1, X_2)$  be a random vector assuming values  $\{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$  such that  $a_1 < a_2$  and  $b_1 < b_2$  with probabilities, respectively,  $p_{11}, p_{12}, p_{21}$  and  $p_{22}$ . In this case,  $F_1(t)$  takes the values 0 if  $t < a_1$ ,  $p_{11} + p_{12}$  if  $a_1 \leq t < a_2$  and 1 otherwise. Similarly,  $F_2(t)$  takes the values 0 if  $t < b_2$ ,  $p_{21} + p_{22}$  if  $b_1 \leq t < b_2$  and 1. Then, the vector  $(F_1(X_1), F_2(X_2))$  takes the values  $\{(p_{11} + p_{12}, p_{21} + p_{22}), (1, p_{21} + p_{22}), (p_{11} + p_{12}, 1), (1, 1)\}$  with probabilities, respectively,  $p_{11}, p_{12}, p_{21}$  and  $p_{22}$ .

**Proposition 7.** [25] Let  $\vec{X}$  and  $\vec{Y}$  be two random vectors with the same copula. Then,  $\vec{X} \leq_{st} \vec{Y}$  if and only if  $X_i \leq_{st} Y_i$  for all  $i \in \{1, \dots, n\}$ .

We end this section by defining a stronger version of the usual stochastic order that will be specially relevant in Section 4.

**Definition 8.** Let  $\vec{X}$  and  $\vec{Y}$  be two random vectors with joint distribution functions  $F$  and  $G$  and marginal distributions  $F_1, \dots, F_n$  and  $G_1, \dots, G_n$ . It is said that  $\vec{X}$  is smaller than or equal to  $\vec{Y}$  in the **same dependence structure stochastic order**, denoted by  $\vec{X} \leq_{sd-st} \vec{Y}$  if  $\vec{X} \leq_{st} \vec{Y}$  and they share the same unique subcopula  $C'$  defined over  $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$  such that

$$F(x_1, \dots, x_n) = C'(F_1(x_1), \dots, F_n(x_n)), \quad G(x_1, \dots, x_n) = C'(G_1(x_1), \dots, G_n(x_n))$$

for any  $\vec{x} \in \mathbb{R}^n$ .

Note that, because of the previous remarks, if  $\vec{X} \leq_{sd-st} \vec{Y}$  then

$$(F_1(X_1), \dots, F_n(X_n)) =_{st} (G_1(Y_1), \dots, G_n(Y_n))$$

**Example 2.** Let  $(X_1, X_2)$  be defined as in Example 1. Consider now a vector  $(Y_1, Y_2)$  similarly defined but assuming values in  $(\tilde{a}_1, \tilde{b}_1), (\tilde{a}_1, \tilde{b}_2), (\tilde{a}_2, \tilde{b}_1)$  and  $(\tilde{a}_2, \tilde{b}_2)$  such that  $\tilde{a}_1 < \tilde{a}_2$  and  $\tilde{b}_1 < \tilde{b}_2$ . Both random vectors share the unique subcopula  $C'$  as considered in Definition 8 and it holds that  $\vec{X} \leq_{sd-st} \vec{Y}$  if and only if  $a_i \leq \tilde{a}_i$  and  $b_i \leq \tilde{b}_i$  for all  $i \in \{1, 2\}$ .

### 2.2. Aggregation of random variables

An aggregation function is typically referred to as a function that summarizes the information of a tuple of values. Formally, given a real interval  $I$  (which can be unbounded), an aggregation function is defined as a monotone increasing function  $A : I^n \rightarrow I$  which satisfies that the infimum and the supremum of its image are, respectively, the infimum and the supremum of the considered interval.

**Definition 9.** [12] Let  $I$  be an interval in the real line  $\mathbb{R}$ . An **aggregation function** is a function  $A : I^n \rightarrow I$  satisfying:

- It is increasing (in each variable).
- The following boundary conditions are fulfilled:

$$\inf_{\vec{x} \in I^n} A(\vec{x}) = \inf I, \quad \sup_{\vec{x} \in I^n} A(\vec{x}) = \sup I.$$

Aggregation functions are widely used in applied problems associated with Data Analysis [3,16,23]. If we follow the usual approach made by Statistics, we can consider the data to be observations of random variables [24]. In this direction, the concept of aggregation of random variables has been introduced in [1]. This type of functions takes a random vector with support over the Cartesian product of an interval and returns a random variable whose image is the same interval. The properties of monotonicity and boundary conditions are defined using a stochastic order. Considering a probability space  $(\Omega, S, P)$ , let us first introduce the following notation:

$$L_I^n(\Omega) = \left\{ \vec{X} : \Omega \rightarrow I^n \mid \vec{X} \text{ is measurable} \right\}.$$

If  $n = 1$ , we will denote  $L_I^1(\Omega)$  just as  $L_I(\Omega)$ .

For the monotonicity, we need to use a stochastic order. To this aim, we can consider a generic stochastic order  $\leq_{SO}$  which is, in general, reflexive, transitive and defined between random vectors of the same dimension  $n$ , for any  $n \in \mathbb{N}$ .

Since stochastic orders are not antisymmetric, the infimum and the supremum of  $L_I$ , needed for the boundary conditions, have to be defined carefully. If  $I$  has an infimum equal to  $a$ , we consider the infimum of  $L_I$  to be the set of random variables having support contained in  $I$  that are minimal with respect to the order  $\leq_{SO}$ . For the stochastic orders  $\leq_{st}, \leq_{icx}$  and  $\leq_{sd-st}$ , mentioned above, for example, this is the set of variables  $X$  such that  $P(X = a) = 1$ . If  $I$  has not a lower bound, then we should apply the same procedure as the one considered in [12], using as infimum a new element  $-\infty$  that is smaller than any other element of  $L_I$ . The definition of the supremum is made analogously.

Then, the definition of an aggregation of random variables is the following

**Definition 10.** [1] Let  $(\Omega, \Sigma, P)$  be a probability space,  $\leq_{SO}$  be a stochastic order and  $I$  be a real non empty interval. An **aggregation function of random variables** (with respect to  $\leq_{SO}$ ) is a function  $A : L_I^n(\Omega) \rightarrow L_I(\Omega)$  which satisfies:

- For any  $\vec{X}, \vec{Y} \in L_I^n(\Omega)$  such that  $\vec{X} \leq_{SO} \vec{Y}$ ,  $A(\vec{X}) \leq_{SO} A(\vec{Y})$ .
- The following boundary conditions with respect to  $\leq_{SO}$  are fulfilled:

$$\inf_{\vec{X} \in L_I^n(\Omega)} A(\vec{X}) = \inf L_I(\Omega), \quad \sup_{\vec{X} \in L_I^n(\Omega)} A(\vec{X}) = \sup L_I(\Omega).$$

Let us discuss the boundary conditions. If  $a$  is the infimum or the supremum of  $I$ , and  $\leq_{SO}$  is anyone of the stochastic orders  $\leq_{st}, \leq_{icx}$  or  $\leq_{sd-st}$  mentioned above, if  $\vec{X} = (a, \dots, a)$  almost surely, then  $A(\vec{X}) =_{a.s} a$ . If  $I$  does not have upper or lower bound, then the image of  $A$  should not have an upper or lower bound with respect to  $\leq_{SO}$ .

In the rest of the paper, we will consider a standard probability space (see [15]) if not specified. Moreover, we will simplify the notation by just writing  $L_I^n$  instead of  $L_I^n(\Omega)$ .

From the wide number of stochastic orders that can be considered in the latter definition, the usual stochastic order which was introduced in Definition 1, has good properties since it allows the composition of measurable aggregation functions and random vectors to be aggregation of random variables. In fact, the following holds.

**Proposition 11.** [1] Let  $I$  be a real interval and let  $A : I^n \rightarrow I$  be a measurable aggregation function. Consider the function  $\hat{A} : L_I^n \rightarrow L_I$  such that for any random vector  $\vec{X} \in L_I^n$ , it holds  $\hat{A}(\vec{X}) = A \circ \vec{X}$ . Then,  $\hat{A}$  is an aggregation function of random variables with respect to the usual stochastic order ( $\leq_{st}$ ).

The function  $\hat{A}$  is referred to as the *aggregation function of random variables induced by the aggregation function  $A$* . Notice that, in general, an aggregation of random variables does not have to be induced by any aggregation function. The most remarkable example is when the construction of the aggregated random variable is made by considering a mixture of the distributions of the components of the initial random vector (see, for example, [7]).

### 2.3. Ordering the arguments in aggregation theory

In deterministic aggregation theory, ordering the arguments is a common way to define aggregation operators. The most prominent example is the Ordered Weighted Averaging (OWA) operator, that makes a convex linear combination of the ordered values.

**Definition 12.** [30] Let  $\vec{w} \in [0, 1]^n$  be a weighting vector such that  $\sum_{i=1}^n w_i = 1$  and  $I$  a real interval. The **Ordered Weighted Averaging Operator**  $OWA_{\vec{w}} : I^n \rightarrow I$  associated to  $\vec{w}$  is defined by:

$$OWA_{\vec{w}}(\vec{x}) = \sum_{k=1}^n w_k x_{\sigma(k)}, \vec{x} \in I^n$$

where  $\sigma$  is a permutation of the  $n$ -tuple  $(1, 2, \dots, n)$  such that  $x_{\sigma(1)} \geq x_{\sigma(2)} \geq \dots \geq x_{\sigma(n)}$ .

The aggregation of random variables induced by an OWA operator is equivalent to consider a convex L-estimator [14], which is a convex linear combination of the order statistics. Recall that, given a random vector  $\vec{X}$ , the order statistics  $X_{(1)}, \dots, X_{(n)}$  are the sorted observations of  $\vec{X}$ , from the smallest to the biggest. We want to clarify that in aggregation theory usually the order is from the biggest to the smallest, although both approaches are equivalent.

The ordering of the arguments is also important in other aggregation functions such as the IOWA operator [31], in which the arguments are sorted by using a secondary vector, or the Choquet and Sugeno integrals [12], in which more complicated computations are made before the ordination.

However, defining ordered aggregation operators become more difficult when dealing with that lack a complete order. The most illustrative example is the aggregation of vectors, where there are cases where it is unclear which vector is the largest (for instance, consider  $\vec{x}_1 = (1, 2)$  and  $\vec{x}_2 = (2, 1)$ ).

Several methods have been defined in the literature in order to avoid the incomparability problem, see [9,17,21]. For the case of multivariate data, we refer to the survey by Pérez-Fernández [22] for more information in this regard. A particular technique introduced in these papers that will be relevant in Section 3 is the use of admissible permutations.

**Definition 13.** [21] Let  $(S, \leq)$  be a partially ordered set and let  $B = \{b_1, \dots, b_n\}$  be a finite sequence in  $S$ . A permutation  $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  is said to be an *admissible permutation* for  $S$  if it holds:

1.  $b_i < b_j \implies \sigma^{-1}(i) < \sigma^{-1}(j)$ .
2. The set  $\{\sigma^{-1}(j) : j \in \{1, \dots, n\} \text{ with } b_j = b\}$  consists of consecutive numbers for any  $b \in B$ .

Notice that, if we have repeated elements in the finite sequence  $B$ , we can treat them as two different elements such that  $b_i = b_j$  in the considered pre-order and the resulting admissible permutations will be the same.

**Example 3.** Consider a pre-ordered set  $\{b_1, b_2, b_3\}$  such that  $b_2 < b_1$  and  $b_2 < b_3$ . Then, two admissible permutations are

$$\begin{array}{l} \sigma_1 : \begin{array}{l} 1 \rightarrow 2 \\ 2 \rightarrow 1 \\ 3 \rightarrow 3 \end{array} \quad \text{and} \quad \sigma_2 : \begin{array}{l} 1 \rightarrow 2 \\ 2 \rightarrow 3 \\ 3 \rightarrow 1 \end{array} \end{array},$$

and no other admissible permutation exists.

In this work a similar problem is presented since stochastic orders are also not complete, thus it is not clear how we can order the random variables before the aggregation. A similar problem has already been considered in the paper by Yager [32], dealing with discrete probability distributions. In there, as an alternative for the stochastic comparison, a point-wise ordination of the distribution functions (called the Probabilistic Exceedance Method) was proposed. We want to remark that our aim is to consider random variables instead of finite discrete probability distributions, but this idea will be revisited in Section 4.

### 3. A first approach: mean ordered weighted averagings

The most naive comparison of the location of two random variables is to compare their expected values. In this direction, we define in this section an aggregation of random variables that initially order the random variables by their mean and then apply a convex linear combination. Since two different random variables can have the same mean, we will consider an average over the set of admissible permutations if there are any ties. In the following, we will consider the order from the smallest to the largest. Given a finite set  $S$ , the number of elements of  $S$  will be denoted as  $\#S$ .

**Definition 14.** Let  $I$  be a real interval and  $\vec{w} \in [0, 1]^n$  a weighting vector such that  $\sum_{i=1}^n w_i = 1$ . Then, a **Mean Ordered Weighted Averaging (MOWA)** is a function  $MOWA_{\vec{w}} : L_I^n \rightarrow L_I$  defined as

$$MOWA_{\vec{w}}(\vec{X}) = \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i X_{\sigma(i)},$$

with  $AP$  being the set of admissible permutations with respect to  $\{E[X_1], \dots, E[X_n]\}$  and the usual order of real numbers.

The last operator is symmetric, i.e. a change in the order of the components of the random vector does not alter the output. Let us give a very simple example of application of the operator:

**Example 4.** Consider a weighting vector  $\vec{w} = (0.2, 0.6, 0.2)$ . Let  $\vec{X} = (X_1, X_2, X_3)$  be a random vector such that their components are independent, and  $P(X_1 = 0) = P(X_1 = 1) = 0.5$ ,  $P(X_2 = 0) = P(X_2 = 2) = 0.5$  and  $P(X_3 = 1) = 1$ . The means of the variables are  $E[X_1] = 0.5$ ,  $E[X_2] = 1$  and  $E[X_3] = 1$ . Therefore, the admissible permutations are

$$\begin{array}{l} \sigma_1 : \begin{array}{l} 1 \rightarrow 1 \\ 2 \rightarrow 2 \\ 3 \rightarrow 3 \end{array} \quad \text{and} \quad \sigma_2 : \begin{array}{l} 1 \rightarrow 1 \\ 2 \rightarrow 3 \\ 3 \rightarrow 2 \end{array} \end{array} .$$

Doing the necessary computations one has,

$$\begin{aligned} MOWA_{\vec{w}}(\vec{X}) &= \frac{1}{2} (0.2X_1 + 0.6X_2 + 0.2X_3) + \frac{1}{2} (0.2X_1 + 0.6X_3 + 0.2X_2) = \\ &= 0.2X_1 + 0.4X_2 + 0.4X_3, \end{aligned}$$

which takes the values 0.4, 0.6, 1.2 and 1.4 with equal probability.

#### 3.1. Monotonicity properties

Although the MOWA seems a natural and easy way to define an ordered operator with respect to the distribution of the random variables, it lacks the monotonicity property. In particular, the only order that can be considered is the one based on the comparison of the expected values. Let us first prove that the mean of the MOWA operator equals to the associated OWA operator applied over the mean vector. We want to remark again that we are considering the order from the smallest one to the greatest one.

**Proposition 15.** Let  $\vec{w} \in [0, 1]^n$  be a weighting vector with  $\sum_{i=1}^n w_i = 1$  and  $\vec{X}$  a random vector. Then,  $E[MOWA_{\vec{w}}(\vec{X})] = OWA_{\vec{w}}(E[\vec{X}])$ .

**Proof.** Considering a particular permutation  $\hat{\sigma} \in AP$  and using the linearity of the expectation, it holds that  $E[MOWA_{\vec{w}}(\vec{X})] = \sum_{i=1}^n w_i E[X_{\hat{\sigma}(i)}] = OWA_{\vec{w}}(E[\vec{X}])$ .  $\square$

**Corollary 16.**  $MOWA_{\vec{w}}$  is an aggregation of random variables with respect to  $\leq_E$  for any weighting vector.

**Proof.** The monotonicity with respect to  $\leq_E$  is a consequence of Proposition 15. The boundary conditions are straightforward to prove recalling that, if  $I$  is bounded, for the  $\leq_E$  order, the set of minimal random variables with support  $I$  is the set of variables  $X$  such that  $P(X = \inf I) = 1$  (and similarly for the supremum). If  $I$  is not bounded, then trivially the expectation is not bounded over the image of  $MOWA_{\vec{w}}$ .  $\square$

The expectation order is a weak stochastic order. Unfortunately, the latter result is not true for the usual stochastic order, as illustrated in the following example.

**Example 5.** In the conditions of Example 4, consider a second random vector  $\vec{Y}$  such that  $Y_1 = X_1$ ,  $Y_2 = X_2$  and  $P(Y_3 = 1.1) = 1$ . Trivially, it holds that  $\vec{X} \leq_{st} \vec{Y}$ . However,

$$MOWA_{\vec{w}}(\vec{Y}) = 0.2Y_1 + 0.6Y_2 + 0.2Y_3,$$

which takes the values 0.22, 0.42, 1.42 and 1.62 with the same probability. If we compute the expected value of the square root, which is an increasing function, we have that  $E\left[\sqrt{MOWA_{\vec{w}}(\vec{X})}\right] \approx 0.921$  and  $E\left[\sqrt{MOWA_{\vec{w}}(\vec{Y})}\right] \approx 0.895$ , thus  $MOWA_{\vec{w}}(\vec{X}) \not\leq_{st} MOWA_{\vec{w}}(\vec{Y})$ .

Another option may be to impose strongest conditions on the initial random vectors. In particular, let us define the *max-min stochastic relation* as  $\vec{X} \leq_{\max\text{-min}} \vec{Y}$  if and only if  $\max \vec{X} \leq_{st} \min \vec{Y}$ , where  $\max \vec{X} = X_{(n)}$  and  $\min \vec{Y} = Y_{(1)}$  are the greatest and smallest order statistics of, respectively,  $\vec{X}$  and  $\vec{Y}$ .

**Proposition 17.** Given a weighting vector  $\vec{w} \in [0, 1]^n$  with  $\sum_{i=1}^n w_i = 1$ ,  $MOWA_{\vec{w}}(\vec{X}) \leq_{st} MOWA_{\vec{w}}(\vec{Y})$  for any pair of random vectors  $\vec{X}$  and  $\vec{Y}$  such that  $\vec{X} \leq_{\min\text{-max}}$ .

**Proof.** Any convex linear combination of  $n$  values takes values between the maximum and the minimum. Therefore, since  $\vec{X} \leq_{\max\text{-min}} \vec{Y}$ , then  $MOWA_{\vec{w}}(\vec{X}) \leq_{a.s.} \max \vec{X} \leq_{st} \min \vec{Y} \leq_{a.s.} MOWA_{\vec{w}}(\vec{Y})$ .  $\square$

We want to remark that the max-min relation is not reflexive (thus is not a stochastic order), since in general one has  $\min \vec{X} <_{st} \max \vec{X}$ . In addition, it is a very strong order and most of random vectors are incomparable in this sense.

### 3.2. An alternative for the multivariate Gaussian distribution

The MOWA has weak monotonicity properties. We devote this section to propose a slight modification of the MOWA operator for Multivariate Gaussian distributions that has a better behavior in this regard. In particular, we aggregate the means of the components and then we normalize the variance in order to allow comparability with respect to the usual stochastic order.

In the following, we will denote as  $\mathcal{G}^{(n)}$  the set of  $n$ -dimensional Gaussian random vectors. For  $n = 1$ , we will use the shortcut  $\mathcal{G}^{(1)} = \mathcal{G}$ . We recall that their distribution is parameterized by a mean vector  $\vec{\mu}$  and a covariance matrix  $\Sigma$ , which is positive semi-definite. As introduced before, the set of positive semi-definite matrices is denoted as  $\mathcal{P}$ . The definition of the operator is as follows.

**Definition 18.** Consider a weighting vector  $\vec{w} \in [0, 1]^n$  such that  $\sum_{i=1}^n w_i = 1$  and a function  $f : \mathcal{P} \rightarrow \mathbb{R}^+$ . The **Gaussian Mean Ordered Weighted Averaging** (GMOWA)  $GMOWA_{\vec{w},f}$  is a function  $GMOWA_{\vec{w},f} : \mathcal{G}^{(n)} \rightarrow \mathcal{G}$  such that for any  $\vec{X} \sim N(\vec{\mu}, \Sigma)$  it holds

$$GMOWA_{\vec{w},f}(\vec{X}) = \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i (\mu_{\sigma(i)} + (X_{\sigma(i)} - \mu_{\sigma(i)}) K),$$

where

$$K = \frac{\sqrt{f(\Sigma)\#AP}}{\sqrt{\text{Var}\left[\sum_{\sigma \in AP} \sum_{i=1}^n w_i X_{\sigma(i)}\right]}}$$

with  $AP$  being the set of admissible permutations with respect to  $\{\mu_1, \dots, \mu_n\}$  and the usual order between real numbers.

The constant  $K$  is defined in order to let the variance of the resulting random variable be only dependent on the covariance matrix of the initial random vector (see subsequent Lemma 19). In general, a reasonable function should preserve the Loewner order, i.e. to be such that if  $\Sigma_1 \leq_{Lo} \Sigma_2$  then  $f(\Sigma_1) \leq f(\Sigma_2)$ . Before studying the properties of GMOWA, let us illustrate this type of operators with a simple example.

**Example 6.** Let us compute the operator when  $\vec{X} \sim N(\vec{\mu}, \Sigma)$  with  $\vec{\mu} = (2, 1)$  and  $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ ,  $f(\Sigma) = \frac{1}{4} \vec{1}^T \Sigma \vec{1}$  and  $\vec{w} = (0.2, 0.8)$ .

Firstly, notice that the only admissible permutation is given by  $\sigma(1) = 2$  and  $\sigma(2) = 1$ , since the mean of  $X_1$  is greater than the mean of  $X_2$ . The variance of the associated linear combination is

$$\text{Var}\left[0.2X_2 + 0.8X_1\right] = 0.2^2 \times 2 + 2 \times 0.2 \times 0.8 + 0.8^2 = 1.04.$$

On the other hand,  $f(\Sigma) = \frac{2+1+1+1}{4} = 1.25$ , so that  $K = \sqrt{\frac{1.25}{1.04}} \approx 1.096$ . Therefore, the GMOWA operator has the expression

$$\begin{aligned} GMOWA_{\vec{w},f}(\vec{X}) &\approx 0.2 \times (1 + (X_2 - 1) \times 1.096) + 0.8 \times (2 + (X_1 - 2) \times 1.096) \\ &\approx 0.219X_2 + 0.877X_1 - 0.173. \end{aligned}$$

Our main objective is to prove that the GMOWA operator satisfies the properties of aggregations of random variables over the set of Gaussian random vectors. We will start with a lemma that gives its distribution.

**Lemma 19.** *The GMOWA  $A_{\vec{w},f}$  fulfill the following properties for any  $\vec{X} \in N(\vec{\mu}, \Sigma)$ :*

1.  $GMOWA_{\vec{w},f}(\vec{X})$  has Gaussian distribution.
2.  $E[GMOWA_{\vec{w},f}(\vec{X})] = OW A_{\vec{w}}(\vec{\mu})$ .
3.  $Var[GMOWA_{\vec{w},f}(\vec{X})] = f(\Sigma)$ .

**Proof.** 1. From the definition, the GMOWA is a linear combination of the components of a multivariate Gaussian random vector, thus it has Gaussian distribution.  
 2. Using well-known properties of the expectation one has

$$\begin{aligned} E \left[ \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i (\mu_{\sigma(i)} + (X_{\sigma(i)} - \mu_{\sigma(i)}) K) \right] &= \\ = \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i \mu_{\sigma(i)} + \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i (E[X_{\sigma(i)}] - \mu_{\sigma(i)}) K &= \\ = \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i \mu_{\sigma(i)} = OW A_{\vec{w}}(\vec{\mu}), \end{aligned}$$

where the last equality holds since for all the admissible permutations the ordered mean vector is the same.

3. Using well-know properties of the variance it follows

$$\begin{aligned} Var \left[ \frac{1}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i (\mu_{\sigma(i)} + (X_{\sigma(i)} - \mu_{\sigma(i)}) K) \right] &= \\ = Var \left[ \frac{K}{\#AP} \sum_{\sigma \in AP} \sum_{i=1}^n w_i X_{\sigma(i)} \right] &= \\ = \frac{K^2}{(\#AP)^2} Var \left[ \sum_{\sigma \in AP} \sum_{i=1}^n w_i X_{\sigma(i)} \right] = f(\Sigma), \end{aligned}$$

where the last equality holds just by replacing the expression of  $K$ .  $\square$

If we compute the moments of the GMOWA associated with Example 6, we can verify that the points 2. and 3. of Lemma 19 hold.

**Example 7.** Considering the latter example, we can compute easily the mean and the variance as

$$\begin{aligned} E [GMOWA_{\vec{w},f}(\vec{X})] &\approx 0.219E[X_2] + 0.877E[X_1] - 0.173 = 1.8 \\ OW A_{\vec{w}}(\vec{\mu}) &= 0.2 \times 1 + 0.8 \times 2 = 1.8 \end{aligned}$$

and

$$Var [GMOWA_{\vec{w},f}(\vec{X})] = 0.219^2 \times 2 + 2 \times 0.219 \times 0.877 + 0.877^2 = 1.25 = f(\Sigma).$$

The next result proves the monotonicity of the GMOWA operator with respect to the usual stochastic order (when considering Gaussian random vectors). Moreover, if  $f$  is increasing with respect to the Loewner order, then it is also monotone with respect to the convex and the increasing convex orders.

**Theorem 20.** *Let  $\vec{X}, \vec{Y}$  two multivariate Gaussian random vectors. Then,  $\vec{X} \leq_{st} \vec{Y}$  implies  $GMOWA_{\vec{w},f}(\vec{X}) \leq_{st} GMOWA_{\vec{w},f}(\vec{Y})$ . In addition, if for any  $\Sigma_1, \Sigma_2 \in \mathcal{P}$  such that  $\Sigma_1 \leq_{Lo} \Sigma_2$  it holds  $f(\Sigma_1) \leq f(\Sigma_2)$ , then  $\vec{X} \leq_{cx} [\leq_{icx}] \vec{Y}$  implies  $GMOWA_{\vec{w},f}(\vec{X}) \leq_{cx} [\leq_{icx}] GMOWA_{\vec{w},f}(\vec{Y})$ .*

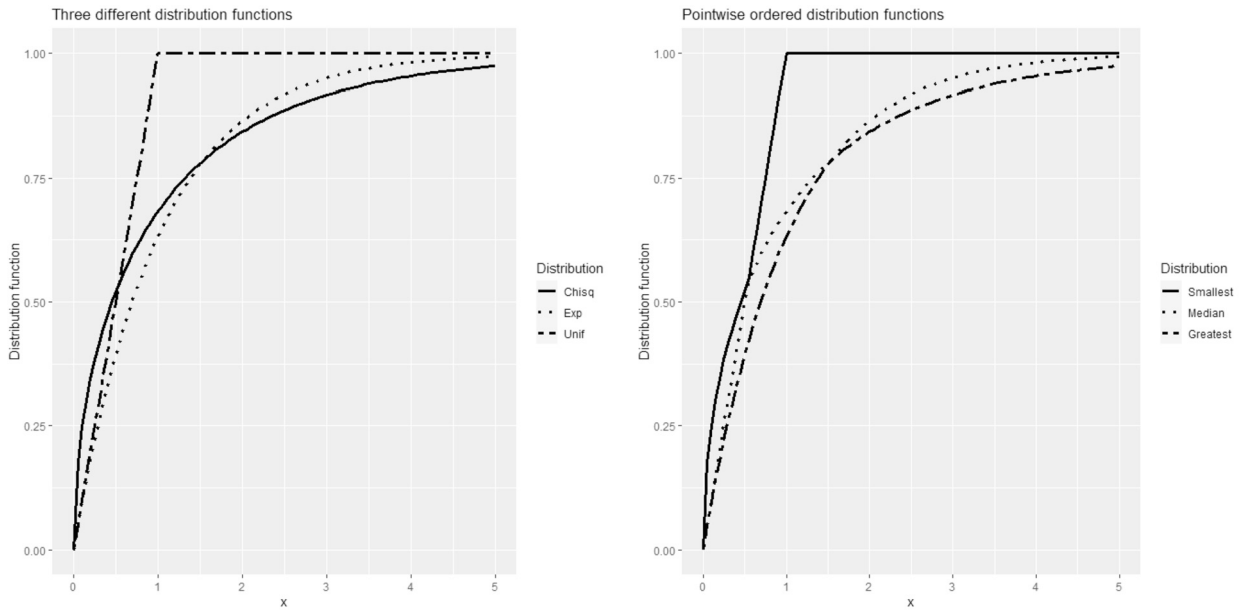


Fig. 1. Three different cumulative distribution functions (left) and the cumulative distribution functions obtained by a pointwise ordination (right).

**Proof.** Consider two random vectors  $\vec{X} \sim N(\vec{\mu}_x, \Sigma_x)$  and  $\vec{Y} \sim N(\vec{\mu}_y, \Sigma_y)$ . Applying Lemma 19,  $GMOW A_{\vec{w},f}(\vec{X})$  has Gaussian distribution with mean  $OW A_{\vec{w}}(\vec{\mu}_x)$  and variance  $f(\Sigma_x)$  (analogously for  $GMOW A_{\vec{w},f}(\vec{Y})$ ). Recalling the conditions for the stochastic comparison of two Multivariate Gaussian distributions given in Theorem 3, the following hold.

- For the usual stochastic order. If  $\vec{X} \leq_{st} \vec{Y}$ , then  $\Sigma_x = \Sigma_y$  and  $\vec{\mu}_x \leq \vec{\mu}_y$ . On the one hand, we have  $\text{Var}[GMOW A_{\vec{w},f}(\vec{X})] = f(\Sigma_x) = f(\Sigma_y) = \text{Var}[GMOW A_{\vec{w},f}(\vec{Y})]$ . The OWA operator is monotone. Then we have  $OW A_{\vec{w}}(\vec{\mu}_x) \leq OW A_{\vec{w}}(\vec{\mu}_y)$  and therefore  $E[GMOW A_{\vec{w},f}(\vec{X})] \leq E[GMOW A_{\vec{w},f}(\vec{Y})]$ .
- For the convex order. If  $\vec{X} \leq_{cx} \vec{Y}$ , then  $\Sigma_x \leq_{Lo} \Sigma_y$  and  $\vec{\mu}_x = \vec{\mu}_y$ . Then,  $\text{Var}[GMOW A_{\vec{w},f}(\vec{X})] = f(\Sigma_x) \leq f(\Sigma_y) = \text{Var}[GMOW A_{\vec{w},f}(\vec{Y})]$ . Trivially, the mean of  $GMOW A_{\vec{w},f}(\vec{X})$  and  $GMOW A_{\vec{w},f}(\vec{Y})$  is the same.
- For the increasing convex order. For the means, proceed as for the usual stochastic order. For the variances, proceed as for the convex order.  $\square$

#### 4. Stochastically ordered aggregation operators

The first approach given in Section 3 has several drawbacks. In the general case, ordering the random variables with respect to their means only allows us to prove the monotonicity for very weak orders under very strong comparison assumptions among the original random vectors. If the variation for the Multivariate Gaussian distribution is considered, we obtain good properties by restricting the aggregation to a particular family of random vectors.

We introduce in this section a set of operators that rely on a transformation, which generates a random vector from any given random vector, ensuring that its components are arranged according to the usual stochastic order. This transformation is based on the idea by Yager in [32], in which several distribution functions are pointwise ordered.

**Example 8.** Consider three random variables, the first having a uniform distribution over the interval  $[0, 1]$ , the second exponentially distributed with parameter  $\lambda = 1$  and the third having a Chi-Squared distribution with 1 degree of freedom. On the left side of Fig. 1, it can be seen that their distribution functions intersect, indicating that they are not ordered according to the usual stochastic order. Ordering pointwise the distribution functions, we obtain three new distributions which are ordered in  $\leq_{st}$ , as shown on the right side of Fig. 1.

In the following, starting with some distribution functions  $F_1, \dots, F_n$ , we will denote as  $F_{[1]}, \dots, F_{[n]}$  the distribution functions obtained as the pointwise ordination of  $F_1, \dots, F_n$ , being  $F_{[1]}$  the greatest distribution function, which is associated with smaller values. They are distribution functions since we can construct them by applying OWA operators (with one of the weights equal to 1 and the rest equal to 0) to the original distribution functions. OWA operators are increasing and continuous [2], therefore we have a composition of increasing functions, which is increasing, and a composition of continuous and right continuous functions, which is right continuous.

With this construction, we obtain a set of ordered distribution functions. However, our objective is to work with random variables and their realizations, not with their distributions. In this direction, we will consider the behavior of random variables by means of the transformation by their distribution functions.

On this aim recall that any random variable  $X$  can be expressed as  $X =_{a.s.} F_X^{-1}(U_X)$ , where  $F_X^{-1}$  is its quantile function and  $U_X$  is a random variable uniform over  $[0, 1]$  (see [20]). Therefore, the outcome of  $X$  can be totally explained by  $U_X$  in the sense that the value of  $X$  is completely determined by the value of  $U_X$ .

For continuous random variables, the uniform distribution is almost surely unique and equals  $F_X(X)$ . In the discrete or mixed case, the decomposition  $X = F_X^{-1}(U_X)$  is not unique, different random variables  $U_X$  can be chosen. Moreover, choosing a uniform random variable will not allow conditional monotonicity (see subsequent Definition 25). A natural election in this case can be, again, to consider  $F_X(X)$  as a random variable that captures the outcome of  $X$ . In this case, the distribution is not longer uniform,  $F_X(X)$  follows a subuniform distribution, see [27,28].

In conclusion, given a set of random variables, we can first build the ordered distributions  $F_{[1]}, \dots, F_{[n]}$  and then assign to them the associated  $F_1(X_1), \dots, F_n(X_n)$  in order to obtain a new random vector whose components are stochastically ordered. The resulting random vector will be called the rearrangement increasing stochastically ordination of the initial one.

**Definition 21.** Let  $\vec{X}$  be a random vector and  $\sigma$  a permutation. Then, the **rearrangement increasing stochastically ordered random vector**, denoted as  $\vec{X}_{so,\sigma}$ , is defined as

$$\vec{X}_{so,\sigma} = \left( F_{[1]}^{-1} \left( F_{\sigma(1)}(X_{\sigma(1)}) \right), \dots, F_{[n]}^{-1} \left( F_{\sigma(n)}(X_{\sigma(n)}) \right) \right),$$

where  $F_{[1]}, \dots, F_{[n]}$  are the distribution functions obtained by the pointwise ordering of  $F_1, \dots, F_n$ , the distribution functions of  $X_1, \dots, X_n$ .

Let us explain in detail the semantic of the permutation  $\sigma$ . If  $\sigma(i) = j$ , we are associating the outcome of  $X_j$  to the distribution  $F_{[i]}^{-1}$  (which is the  $i$ -th smallest). It allows  $\vec{X}_{so,\sigma}$  to be totally determined by  $\vec{X}$  in the sense that the conditional distribution of  $\vec{X}_{so,\sigma}$  given a value of  $\vec{X}$  is degenerate. This will be relevant in applied problems (see Section 5) since any observation of the random vector will have an associated value of the rearrangement increasing stochastically ordered random vector.

A Stochastically Ordered Aggregation is a function that takes a random vectors and returns a random variable that is obtained as the composition of a usual aggregation function and the rearrangement increasing stochastically ordered random vector. Formally, the following is its definition.

**Definition 22.** Let  $I$  be a real interval,  $A : I^n \rightarrow I$  a measurable aggregation function and  $\sigma$  a permutation. Then, the **Stochastically Ordered Aggregation (SOA)** is a function  $SOA_{A,\sigma} : L_I^n \rightarrow L_I$  such that for any random vector  $\vec{X} \in L_I^n$ ,  $SOA_{A,\sigma}(\vec{X}) = A \circ \vec{X}_{so,\sigma}$ .

We can see the Stochastically Ordered Aggregations as induced aggregations applied over rearrangement increasing stochastically ordered random vectors. Before studying its properties, let us illustrate its definition with an example.

**Example 9.** Let  $\vec{X} = (X_1, X_2)$  be a random vector with independent components such that  $X_1 \sim Exp(1.5)$ ,  $X_2 \sim U[0, 1]$ . It is easy to see that for  $F_1$  and  $F_2$  it holds  $F_2(t) \leq F_1(t)$  if  $t \leq t_0$  and  $F_1(t) \leq F_2(t)$  if  $t \geq t_0$ , with  $t_0 \approx 0.583$ . Considering the permutation  $\sigma$  such that  $\sigma(1) = 2$  and  $\sigma(2) = 1$  and  $A$  the arithmetic mean, we have the following expression for the  $SOA_{A,\sigma}(\vec{X})$  operator:

$$SOA_{A,\sigma}(\vec{X}) = \begin{cases} \frac{1}{2}(X_2 + X_1), & \text{if } 1 - e^{-1.5X_1} > t_0, X_2 \leq t_0 \\ \frac{1}{2}(X_2 + (1 - e^{-1.5X_1})), & \text{if } 1 - e^{-1.5X_1} \leq t_0, X_2 \leq t_0 \\ \frac{1}{2}(-\ln(1 - X_2) + X_1), & \text{if } 1 - e^{-1.5X_1} > t_0, X_2 > t_0 \\ \frac{1}{2}(-\ln(1 - X_2) + (1 - e^{-1.5X_1})), & \text{if } 1 - e^{-1.5X_1} \leq t_0, X_2 > t_0 \end{cases}$$

#### 4.1. Monotonicity properties

The main desirable property of our operator is the monotonicity with respect a stochastic order. We will start showing a preliminary result regarding the rearrangement increasing stochastically ordered random vector. For it, recall the definition of  $\leq_{sd-st}$  order provided in Definition 8.

**Theorem 23.** Let  $\vec{X}$  and  $\vec{Y}$  be two random vectors such that  $\vec{X} \leq_{sd-st} \vec{Y}$  and let  $\sigma$  be any permutation. Then,  $\vec{X}_{so,\sigma} \leq_{st} \vec{Y}_{so,\sigma}$ .

**Proof.** Denote as  $F_1, \dots, F_n$  and  $G_1, \dots, G_n$  the marginal distribution functions of  $\vec{X}$  and  $\vec{Y}$  respectively. Since  $\vec{X} \leq_{sd-st} \vec{Y}$  implies  $\vec{X} \leq_{st} \vec{Y}$  and therefore  $X_i \leq_{st} Y_i$  for any  $i = \{1, \dots, n\}$ , we have that  $F_i(t) \geq G_i(t)$  for any  $i = \{1, \dots, n\}$  and  $t \in \mathbb{R}$ . Then, it holds that  $F_{[i]}(t) \geq G_{[i]}(t)$  for any  $i = \{1, \dots, n\}$  and  $t \in \mathbb{R}$  and also  $F_{[i]}^{-1}(t) \leq G_{[i]}^{-1}(t)$  for any  $i = \{1, \dots, n\}$  and  $t \in [0, 1]$

As remarked after Definition 8,  $(F_1(X_1), \dots, F_n(X_n))$  and  $(G_1(Y_1), \dots, G_n(Y_n))$  have the same distribution. Consider a random vector  $\vec{Z}$  with the same distribution as  $(F_1(X_1), \dots, F_n(X_n))$  and  $(G_1(Y_1), \dots, G_n(Y_n))$ . Then,

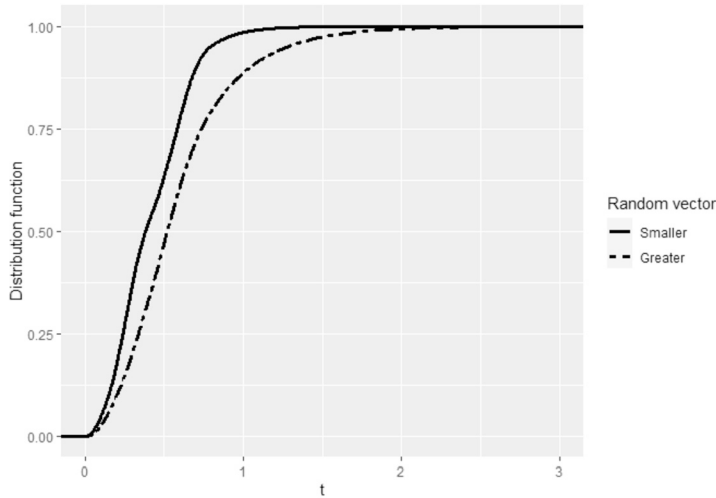


Fig. 2. Simulated cumulative distribution functions of a Stochastically Ordered Aggregation of two ordered random vectors with sample size  $n = 10^6$ .

$$\begin{aligned} \vec{X}_{so,\sigma} &=_{st} \left( F_{[1]}^{-1}(Z_{\sigma(1)}), \dots, F_{[n]}^{-1}(Z_{\sigma(n)}) \right) \leq_{a.s.} \\ &\leq_{a.s.} \left( G_{[1]}^{-1}(Z_{\sigma(1)}), \dots, G_{[n]}^{-1}(Z_{\sigma(n)}) \right) =_{st} \vec{Y}_{so,\sigma}, \end{aligned}$$

which implies the usual stochastic order by using Theorem 2.  $\square$

As a consequence of the latter result, it can be proved that the Stochastically Ordered Aggregations are aggregation of random variables with respect to the same dependence structure stochastic order.

**Corollary 24.** Any Stochastically Ordered Aggregation  $SOA_{A,\sigma}$  is an aggregation of random variables with respect to  $\leq_{sd-st}$  for any aggregation function  $A : I^n \rightarrow I$  and any permutation  $\sigma$ .

**Proof.** Let  $\vec{X} \leq_{sd-st} \vec{Y}$ . Using Theorem 23, we have that  $\vec{X}_{so,\sigma} \leq_{st} \vec{Y}_{so,\sigma}$  for any permutation  $\sigma$ . Since  $SOA_{A,\sigma}(\vec{X}) = A \circ \vec{X}_{so,\sigma}$  and  $SOA_{A,\sigma}(\vec{Y}) = A \circ \vec{Y}_{so,\sigma}$ , applying Proposition 11, it is concluded that  $SOA_{A,\sigma}(\vec{X}) \leq_{st} SOA_{A,\sigma}(\vec{Y})$ . The usual stochastic order and the same copula usual stochastic order are equivalent for random variables, thus the monotonicity holds. The boundary conditions are straightforward to prove.  $\square$

**Example 10.** Consider the same conditions as in Example 9 and let  $\vec{Y}$  be a random vector with independent components such that  $Y_1 \sim U[0.5, 1.5]$  and  $Y_2 \sim Exp(1)$ . Trivially,  $\vec{X} \leq_{st} \vec{Y}$  and both random vectors have the same and unique copula, thus  $\vec{X} \leq_{sd-st} \vec{Y}$ . If we simulate the distribution of  $SOA_{A,\sigma}(\vec{X})$  and  $SOA_{A,\sigma}(\vec{Y})$ , plotting their empirical distributions one can observe that they are ordered, see Fig. 2. Therefore, we can see that  $SOA_{A,\sigma}(\vec{X}) \leq_{st} SOA_{A,\sigma}(\vec{Y})$ .

Another interesting monotonicity property that is specially relevant in applied problems, see Section 5, is the conditional monotonicity. Given a fixed random vector  $\vec{X}$ , we wonder if  $SOA(\vec{X}, A, \sigma)$  is degenerate and, moreover, if it increases when the value of  $\vec{X}$  increases.

**Definition 25.** Let  $I$  be a real interval and  $A : L_I^n \rightarrow L_I$  an aggregation of random variables. Then,  $A$  is said to be **conditional increasing** if, for any  $\vec{X} \in L_I^n$  with support  $S$ :

- $[A(\vec{X}) \mid \vec{X} = \vec{x}]$  has degenerate distribution for any  $\vec{x} \in S$ .
- The function  $\hat{A} : S \rightarrow I$  defined as  $\hat{A}(\vec{x}) = \lambda$ , being  $\lambda$  the value such that  $P([A(\vec{X}) \mid \vec{X} = \vec{x}] = \lambda) = 1$ , is an increasing function.

We can see this property as a sort of comonotonicity property of the initial random vector and the output random variable. Another interpretation is that conditional increasing aggregations applied over observations of the random vectors are increasing functions from vectors to real numbers. The *MOWA* and *GOWA* operators are trivially conditional increasing. Let us prove the same statement for the *SOA* operator.

**Proposition 26.** Any Stochastically Ordered Aggregation  $SOA_{A,\sigma}$  is conditional increasing.

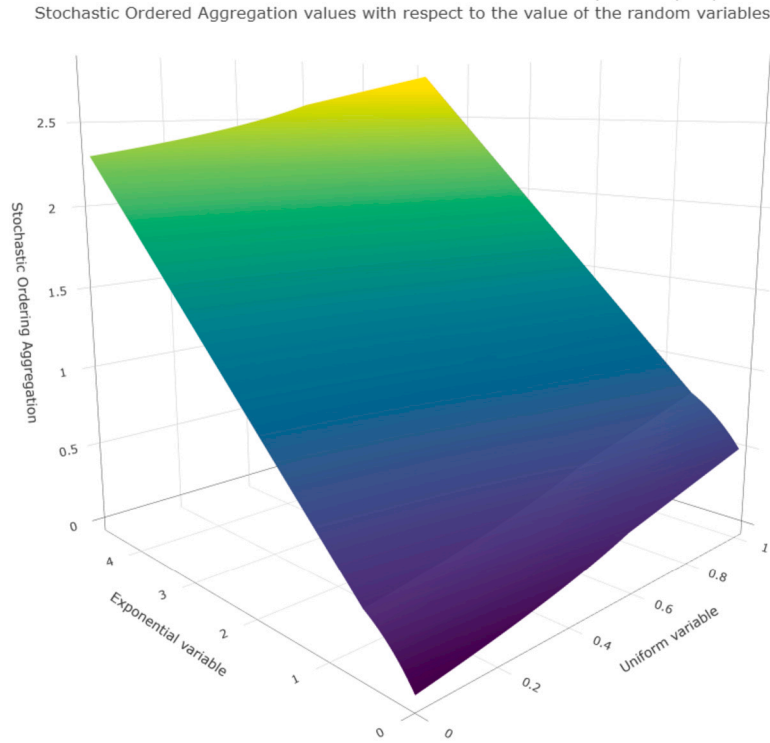


Fig. 3. Representation of the function  $\hat{A}$  as defined in Definition 25 for Example 9.

**Proof.** Consider a Stochastically Ordered Aggregation  $SOA_{A,\sigma} : L_I^n \rightarrow L_I$  and a random vector  $\vec{X} \in L_I^n$  with support  $S$ . For any  $\vec{x} \in S$ , we have that

$$SOA(\vec{x}, A, \sigma) = A \left( F_{[1]}^{-1} (F_{\sigma(1)}(x_{\sigma(1)})), \dots, F_{[n]}^{-1} (F_{\sigma(n)}(x_{\sigma(n)})) \right).$$

From the latter expression, we have that  $[SOA(\vec{X}, A, \sigma) \mid \vec{X} = \vec{x}]$  has degenerate distribution for any  $\vec{x} \in S$ . In addition, the function  $\hat{A} : S \rightarrow I$  of Definition 25 is increasing since it is a composition of the increasing functions  $F_{[1]}^{-1}, \dots, F_{[n]}^{-1}, F_1, \dots, F_n$  and  $A$ .  $\square$

Let us illustrate this property with an example.

**Example 11.** Consider the SOA operator considered in Example 9. If we represent the function  $\hat{A}$  as defined in Definition 25, we obtain the surface represented in Fig. 3. As it can be seen, the function is increasing, thus the value of the Stochastically Ordered Operator increases with the values assumed by the aggregated variables. Note also that the boundary condition on  $(0, 0)$  is satisfied.

Leaving aside the monotonicity properties, we want to end this section by stating a lemma that links the Stochastically Ordered Operators with aggregation functions applied over the order statistics.

**Lemma 27.** Let  $\vec{X}$  be a random vector with continuous comonotone components. Then  $SOA_{A,\sigma}(\vec{X}) =_{a.s.} A \circ (X_{(1)}, \dots, X_{(n)})$ .

**Proof.** Since  $\vec{X}$  has comonotone components, then there exists a uniform random variable  $U$  such that

$$\vec{X} =_{a.s.} (F_1(U), \dots, F_n(U)).$$

By construction, we have that the random vector  $(F_{[1]}^{-1}(U), \dots, F_{[n]}^{-1}(U))$  is a random vector in which, with probability one, the components are ordered. Therefore, we have that

$$(X_{(1)}, \dots, X_{(n)}) =_{as} (F_{[1]}^{-1}(U), \dots, F_{[n]}^{-1}(U)) =_{a.s.} \vec{X}_{so,\sigma}$$

for any permutation  $\sigma$ , since the components are continuous. Then, the result holds by noting that  $SOA_{A,\sigma}(\vec{X}) = A \circ \vec{X}_{so,\sigma}$ .  $\square$

### 4.2. The stochastically ordered weighting averaging

This section is devoted to the study of a particular case of Stochastically Ordered Aggregations for which the considered aggregation function is a weighted averaging. From a theoretical point of view, this particular choice is interesting because additional theoretical properties can be proved. From a practical one, it is a family of aggregations of random variables that are parameterized by a permutation and a weighting vector, which can be optimized easily when working with data (see Section 5). Let us start by defining the concept itself.

**Definition 28.** Let  $\vec{w} \in [0, 1]^n$  be a weighting vector such that  $\sum_{i=1}^n w_i = 1$  and  $I$  a real interval. The **Stochastic Ordered Weighting Averaging (SOWA)** Operator  $SOWA_{\vec{w},\sigma} : L_I^n \rightarrow L_I$  associated to  $\vec{w}$  is defined by

$$SOWA_{\vec{w},\sigma}(\vec{X}) = \sum_{k=1}^n w_k \left( \vec{X}_{s_{\sigma(k)}} \right)_k, \quad \vec{X} \in L_I^n.$$

This operator is specially relevant, since it allows to give more or less importance to each of the distributions  $F_{[1]}, \dots, F_{[n]}$ . For instance, we can give small weights to the extreme distributions, giving less importance to them. This can be seen as a similar procedure as when using the OWA operator with greater middle weights (see [10]), but instead of reducing the impact of extreme observations (outliers), we are minimizing the impact of extreme distributions.

In the following result we prove several linearity properties of the SOWA operator.

**Proposition 29.** The  $SOWA_{\vec{w},\sigma}(\vec{X})$  operator fulfills the following properties for any weighting vector  $\vec{w}$  and permutation  $\sigma$ :

- $SOWA_{\vec{w},\sigma}(\lambda \vec{X}) = \lambda SOWA_{\vec{w},\sigma}(\vec{X})$  for any random vector  $\vec{X}$  and  $\lambda \in \mathbb{R}^+$ .
- $SOWA_{\vec{w},\sigma}(\vec{X} + \lambda \vec{1}) = SOWA_{\vec{w},\sigma}(\vec{X}) + \lambda$  for any random vector  $\vec{X}$  and  $\lambda \in \mathbb{R}$ .
- $SOWA_{\vec{w},\sigma}(\vec{X} + \vec{Y}) = SOWA_{\vec{w},\sigma}(\vec{X}) + SOWA_{\vec{w},\sigma}(\vec{Y})$  for any pair of random vectors such that there exists a permutation  $\hat{\sigma}$  for which  $X_{\hat{\sigma}(1)} \leq_{st} \dots \leq_{st} X_{\hat{\sigma}(n)}$  and  $Y_{\hat{\sigma}(1)} \leq_{st} \dots \leq_{st} Y_{\hat{\sigma}(n)}$  are satisfied and the random variables  $X_i$  and  $Y_i$  are comonotone for any  $i \in \{1, \dots, n\}$ .

**Proof.** The first two statements are straightforward to prove from the definition of SOWA. For the third one, let us denote as  $F_1, \dots, F_n, G_1, \dots, G_n$  and  $H_1, \dots, H_n$  the distribution functions of, respectively,  $\vec{X}, \vec{Y}$  and  $\vec{X} + \vec{Y}$ . Note that, by hypothesis,  $X_i$  and  $Y_i$ , and therefore  $X_i + Y_i$  with  $X_i + Y_i$  and  $Y_i$  with  $X_i + Y_i$ , are comonotone for any  $i \in \{1, \dots, n\}$ . Thus, there exists a standard uniform random variable such that  $X_i =_{a.s.} F_i^{-1}(U_i)$ ,  $Y_i =_{a.s.} G_i^{-1}(U_i)$  and  $X_i + Y_i =_{a.s.} H_i^{-1}(U_i)$  for any  $i \in \{1, \dots, n\}$ . From the equivalence between the two expressions for  $X_i + Y_i$ , it holds that  $H_i^{-1}(U_i) =_{a.s.} F_i^{-1}(U_i) + G_i^{-1}(U_i)$  and  $H_i^{-1} = F_i^{-1} + G_i^{-1}$  (almost everywhere) for any  $i \in \{1, \dots, n\}$ .

Keeping in mind the latter considerations about comonotonicity of the marginals, consider the random vector  $\vec{U}$  with standard uniform marginal distributions such that:

$$\begin{aligned} \vec{U} &=_{a.s.} (X_1, \dots, X_n) =_{a.s.} (G_1(Y_1), \dots, G_n(Y_n)) =_{a.s.} \\ &=_{a.s.} (H_1(X_1 + Y_1), \dots, H_n(X_n + Y_n)). \end{aligned}$$

In addition, since the components of  $\vec{X}$  and  $\vec{Y}$  are ordered in the same order according to  $\leq_{st}$ , we have  $H_{[i]}^{-1} = F_{[i]}^{-1} + G_{[i]}^{-1}$  for any  $i \in \{1, \dots, n\}$ . Then

$$\begin{aligned} SOWA_{\vec{w},\sigma}(\vec{X} + \vec{Y}) &= \sum_{i=1}^n w_i H_{[i]}^{-1}(U_{\sigma(i)}) = \\ &= \sum_{i=1}^n w_i \left( F_{[i]}^{-1}(U_{\sigma(i)}) + G_{[i]}^{-1}(U_{\sigma(i)}) \right) = SOWA_{\vec{w},\sigma}(\vec{X}) + SOWA_{\vec{w},\sigma}(\vec{Y}). \quad \square \end{aligned}$$

In addition, for continuous random vectors, its expected value has the same value as the one defined by the linear combination of the order statistics.

**Proposition 30.** Let  $\vec{X}$  be a random vector with continuous components. Then  $E[SOWA_{\vec{w},\sigma}(\vec{X})] = \sum_{i=1}^n w_i E[X_{(i)}]$ .

**Proof.** Noting that the components are continuous, denote as  $U_1, \dots, U_n$  the standard uniform random variables such that  $U_i = F_i(X_i)$  with  $i \in \{1, \dots, n\}$ . Then,

$$E[SOWA_{\vec{w},\sigma}(\vec{X})] = E \left[ \sum_{i=1}^n w_i F_{(i)}^{-1}(U_{\sigma(i)}) \right] = \sum_{i=1}^n w_i E \left[ F_{(i)}^{-1}(U_{\sigma(i)}) \right] =$$

$$= \sum_{i=1}^n w_i \int_0^1 F_{(i)}^{-1}(t) dt = \int_0^1 \left( \sum_{i=1}^n w_i F_{(i)}^{-1}(t) \right) dt.$$

This expression does not depend on the copula of the random vector. The result holds since, using Lemma 27, for the case of comonotone components we have  $SOWA_{\vec{w},\sigma}(\vec{X}) = a.s. \sum_{i=1}^n w_i X_{(i)}$ .  $\square$

We want to remark that the expression  $\sum_{i=1}^n w_i X_{(i)}$  can be seen as the aggregation of random variables induced by an OWA operator (but ordering from the smallest to the greatest).

### 5. A practical example regarding temperature prediction

We devote this section to illustrate the advantages of the use of the presented aggregation operators in predictions problems. In the following, we will consider a scenario in which a random vector  $\vec{X}$  is used to predict the value of a random variable  $Y$ . The prediction will be denoted as  $\hat{Y}$ . The first three alternatives will be the following,

- (1) WAM:  $\hat{Y} = \sum_{i=1}^n w_i X_i$ ,
- (2) OWA:  $\hat{Y} = \sum_{i=1}^n w_i X_{(i)}$ ,
- (3) SOWA:  $\hat{Y} = SOWA_{\vec{w},\sigma}(\vec{X})$ ,

which are convex linear combination of, respectively, the random variables (WAM), the order statistics (OWA) and the components of the rearrangement increasing stochastically ordered random vector (SOWA). Leaving aside our proposal, we have chosen these alternatives since they are the most used aggregation functions that involve, as well as the SOWA operator, a weight vector and, in the case of the OWA operator, a permutation of the observed values, therefore we can compare the resulting parameters.

Notice that the use of a MOWA operator in this context is equivalent to the use of a weighted mean. The possible permutation of the components of the initial random vector can be undone by permuting in the same way the weighting vector.

In particular, we have used the database Appliances Energy Prediction, which can be found in [4]. The considered data consists in the value of the temperature (in Celsius) in different locations of a house, measured in 19735 different times. We refer to [5] for more details about the dataset. If we dismiss the values associated with outside locations (since its temperature is quite different of the temperature inside the house), we have the temperatures for the *kitchen*, *living room*, *laundry room*, *bathroom*, *teenager room*, *ironing room*, *parents room* and *office room*. Our aim is to predict the temperature in the *office room* using the values in the rest of indoors locations, by considering the models (1-3). The variables are ordered in the vector in the order in which they have been enumerated above.

We have carried on two studies, one of them using a fixed training and test samples and focused on the semantics and interpretation of the optimal parameters and a second one using cross-validation and focused on the performance of the different models and including more involved prediction methods.

In the first one, we divided the dataset in a 75%, the training dataset, which is used to optimize the parameters of the models and the test dataset, which is reserved for testing the behavior of the models. The optimization criterion is the usual minimization of the Mean Squared Error. For the models (1) and (2), this process is quite simple, the feasible region is the set of weighting vectors.

For the case of the Stochastically Ordered Weighting Averaging, we first need to build the rearrangement increasing stochastically ordered random vector. Since we do not know the population distributions of the variables, we have used empirical distribution functions (see [24]) of the random variables as the  $F_1, \dots, F_n$  in Definition 21. Then, the optimization is made in the set of all possible permutations  $\sigma$  (which are  $7! = 5040$ ) and the set of weighting vectors. The obtained optimal parameters for all the models are the following:

$$\begin{aligned} \vec{w}_{(1)} &= (0.36, 0.09, 0.09, 0.04, 0, 0.20, 0.21), \\ \vec{w}_{(2)} &= (0.02, 0.20, 0.24, 0.017, 0.43, 0.09, 0), \\ \vec{w}_{(3)} &= (0.04, 0.25, 0.12, 0.27, 0.17, 0.16, 0), \\ \sigma &= (4, 7, 2, 6, 1, 3, 5). \end{aligned}$$

For the weighted mean, the variables with more importance are *kitchen*, *parents room* and *ironing room*. In the combination of the order statistics, it can be seen that the extreme values, the maximum and the minimum, do not have much importance, since the associated weights are almost 0. Something similar happens in the case of the SOWA operator, the extreme distributions are given less importance. Note also that, as can be seen in the optimal permutation, the outcome of the temperature in the *teenager room* is associated with the greatest distribution, and therefore, with a null weight. This coincides to the weight associated to the *teenager room* in model (1).

The Mean Squared Error, Mean Absolute error and Percent error (see [19]) of the three models in the test sample are provided in Table 1. As it can be seen, the SOWA has a better behavior than the other alternatives.

Latter study is useful for illustrating the semantic of the parameters of the SOWA operator in a applied example, as well as the relation of them with the weights of the WAM and OWA operators. However, a deeper study of the accuracy of the proposed

**Table 1**

Mean Squared Error (MSE), Mean Absolute Error (MAE) and Percent Error (PE) for the considered prediction models in the test sample.

Model	MSE	MAE	PE (%)
WAM	0.28	0.44	1.93
OWA	0.31	0.46	1.97
SOWA	0.22	0.38	1.66

**Table 2**

Average Mean Squared Error (MSE), Mean Absolute Error (MAE) and Percent Error (PE) for the considered prediction models in the cross-validation procedure.

Model	MSE	MAE	PE (%)
WAM	0.86	0.73	3.61
OWA	0.67	0.65	3.17
SOWA	0.58	0.58	2.83
NN	0.72	0.67	3.24
LR	0.62	0.61	3.00
RF	0.75	0.69	3.39

method should be done, since the results are sensitive to the election of the training and test samples and more involved prediction models should be considered.

Therefore, the sample has been divided in 10 blocks of the same size and a cross-validation procedure (see [6]) has been applied. In particular, 9 of the blocks have been used as the training sample and the remaining one as the test sample, for each possible combination. Moreover, three additional prediction models have been considered as follows.

- (4) NN: A neural network consisting on a input layer of 7 neurons, a hidden layer of 3 neurons and an output layer of 1 neuron has been considered. As algorithm, we have chosen resilient backpropagation with weight backtracking. MSE was chosen as the error function and the logistic function as the activation function. The learning rate has been considered as 0.5, the stopping criterion threshold as 0.5 and 10 repetitions of the training. We have used the R package *neuralnet*, we refer to [13] for more information.
- (5) LR: The usual linear regression procedure, see Chapter 12 in [24].
- (6) RF: A random forest consisting of 100 trees with maximum depth. For the implementation, the R package *randomForest* has been used, see [18].

The average of the errors can be found in Table 2. It can be seen that the average errors are greater than in the first study and that the SOWA operator has the better behavior among all alternatives.

The *SOWA* operator, as constructed in this section, only can take values between the minimum and the maximum of the inputs in the training data. Therefore, it is adequate to use it only when the predicted variables have the same order of magnitude as the explanatory variables.

The here-presented results are just particular example that serves as an illustration of the potential use of the *SOWA* operator. In general, for other databases or for more involved prediction models, one cannot expect the *SOWA* operator to have the best behavior, in particular with respect to neural networks and random forests.

Note that other aggregations, not only weighted averages, may be considered in prediction problems. In those cases, the Stochastically Ordered Aggregation associated to the particular aggregation can be always considered. We want also remark that if the dimension of the aggregated vectors is too big, the optimization over the set of permutations can be unaffordable.

## 6. Conclusions

The definition and properties of aggregation of random variables that consider a stochastic comparison of the input random variables has been studied. This approach permits to consider the whole distribution, not just the value of a particular observation, in the ordination of the arguments of the aggregation.

Firstly, the Mean Ordered Weighted Averaging, which orders the variables considering their means are defined. The value of its mean has been obtained and the monotonicity with respect to the expectation order and the min-max stochastic order has been proved.

A variation for multivariate Gaussian distribution has been also explored. In this case, the operator is monotonic with respect to the usual stochastic order, and, under additional conditions, is also monotonic with respect to the convex and increasing convex orders. Its distribution has been derived.

Following a similar idea as Yager in [32], the Stochastically Ordered Aggregations have been defined. The monotonicity with respect the same dependence structure stochastic order has been proved, as well as the conditional increasing property. For the particular case of the Stochastically Ordered Weighting Averaging, some linear properties and its expectation are provided.

Finally, the behavior of the Stochastically Ordered Weighted Averaging is compared with the Weighted Averaging and the Ordered Weighted Averaging in a practical example involving the prediction of the temperature in a house. The here-proposed method has a better behavior in all the considered error measures.

We expect this work to open up the possibility of more definitions of aggregations that take into account not only particular observations of variables but also their probabilistic behavior by means, for instance, of their moments or their cumulative distribution functions.

## Funding details

J. Baz is partially supported by Programa Severo Ochoa of Principality of Asturias (BP21042). F. Pellerey carried out this study within the FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) Missione 4 Componente 2, Investimento 1.3 D.D. 1555 11/10/2022, PE00000013). J. Baz, S. Montes and I. Díaz have been supported by the Ministry of Science and Innovation (PDI2022-139886NB-I00).

## CRedit authorship contribution statement

**Juan Baz:** Conceptualization, Investigation, Methodology, Writing – original draft. **Franco Pellerey:** Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. **Irene Díaz:** Conceptualization, Supervision, Writing – review & editing. **Susana Montes:** Conceptualization, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data can be found online. The source is cited in the manuscript.

## References

- [1] J. Baz, I. Díaz, S. Montes, The choice of an appropriate stochastic order to aggregate random variables, in: *Building Bridges Between Soft and Statistical Methodologies for Data Science*, Springer, 2022, pp. 40–47.
- [2] G. Beliakov, S. James, J. Wu, *Discrete Fuzzy Measures*, Springer, 2020.
- [3] G. Beliakov, A. Pradera, T. Calvo, et al., *Aggregation Functions: A Guide for Practitioners*, vol. 221, Springer, 2007.
- [4] L.M. Candanedo, V. Feldheim, D. Deramaix, Appliances energy prediction dataset, Online; <https://www.kaggle.com/datasets/loveall/appliances-energy-prediction>. (Accessed 1 May 2023), 2017.
- [5] L.M. Candanedo, V. Feldheim, D. Deramaix, Data driven prediction models of energy use of appliances in a low-energy house, *Energy Build.* 140 (2017) 81–97.
- [6] G.C. Cawley, N.L. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, *J. Mach. Learn. Res.* 11 (2010) 2079–2107.
- [7] S. Chandra, On the mixtures of probability distributions, *Scand. J. Stat.* (1977) 105–112.
- [8] E. de Amo, M.D. Carrillo, J. Fernández-Sánchez, Characterization of all copulas associated with non-continuous random variables, *Fuzzy Sets Syst.* 191 (2012) 103–112.
- [9] M. Ferrero-Jaurrieta, L. Horanská, J. Lafuente, R. Mesiar, G.P. Dimuro, Z. Takáč, M. Gómez, J. Fernández, H. Bustince, Degree of totalness: how to choose the best admissible permutation for vector fuzzy integration, in: *Fuzzy Sets and Systems*, 2023.
- [10] D. García-Zamora, Á. Labella, R.M. Rodríguez, L. Martínez, Symmetric weights for OWA operators prioritizing intermediate values. The EVR-OWA operator, *Inf. Sci.* 584 (2022) 583–602.
- [11] C. Genest, J. Nešlehová, A primer on copulas for count data, *ASTIN Bulletin: The Journal of the IAA* 37 (2) (2007) 475–515.
- [12] M. Grabisch, J.-L. Marichal, R. Mesiar, E. Pap, *Aggregation Functions*, vol. 127, Cambridge University Press, 2009.
- [13] F. Günther, S. Fritsch, Neuralnet: training of neural networks, *R J.* 2 (1) (2010) 30.
- [14] J. Hosking, L-estimation, in: *Order Statistics: Applications*, in: *Handbook of Statistics*, vol. 17, Elsevier, Amsterdam, Netherlands, 1998, pp. 215–235.
- [15] K. Itô, et al., *An Introduction to Probability Theory*, Cambridge University Press, 1984.
- [16] S. James, *An Introduction to Data Analysis Using Aggregation Functions in R*, Springer, 2016.
- [17] L. Jin, R. Mesiar, R. Yager, Ordered weighted averaging aggregation on convex poset, *IEEE Trans. Fuzzy Syst.* 27 (3) (2019) 612–617.
- [18] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, *R News* 2 (3) (2002) 18–22.
- [19] D.C. Montgomery, C.L. Jennings, M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons, 2015.
- [20] R.B. Nelsen, *An Introduction to Copulas*, Springer Science & Business Media, 2007.
- [21] D. Paternain, L. Jin, R. Mesiar, L. Vavříková, H. Bustince, OWA operators based on admissible permutations, in: *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 2019, pp. 1–5.
- [22] R. Pérez-Fernández, Multivariate OWA functions, *Fuzzy Sets Syst.* (2022).
- [23] N. Pollesch, V.H. Dale, Applications of aggregation theory to sustainability assessment, *Ecol. Econ.* 114 (2015) 117–127.

- [24] V.K. Rohatgi, A.M.E. Saleh, *An Introduction to Probability and Statistics*, John Wiley & Sons, 2015.
- [25] M. Shaked, J.G. Shanthikumar, *Stochastic Orders*, Springer, 2007.
- [26] M. Sklar, Fonctions de repartition an dimensions et leurs marges, *Publ. Inst. Stat. Univ. Paris* 8 (1959) 229–231.
- [27] S. Tasena, On a distribution form of subcopulas, *Int. J. Approx. Reason.* 128 (2021) 1–19.
- [28] S. Tasena, Central limit theorem for subcopulas under the Manhattan distance, *J. Math. Anal. Appl.* 509 (2) (2022) 126007.
- [29] J.J. Van De Gronde, J.B. Roerdink, Frames, the Loewner order and eigendecomposition for morphological operators on tensor fields, *Pattern Recognit. Lett.* 47 (2014) 40–49.
- [30] R.R. Yager, Families of OWA operators, *Fuzzy Sets Syst.* 59 (2) (1993) 125–148.
- [31] R.R. Yager, Induced aggregation operators, *Fuzzy Sets Syst.* 137 (1) (2003) 59–69.
- [32] R.R. Yager, On the OWA aggregation with probabilistic inputs, *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 23 (Suppl. 1) (2015) 143–162.