Taxonomy of aggregations of random variables *

Juan Baz[1][0000−0003−1142−6077], Gleb Beliakov[2][0000−0002−9841−5292], Irene Díaz[3][0000−0002−3024−6605], and Susana Montes[1][0000−0002−4701−2207]

[1] Department of Statistics and I.O. and Didactic of Mathematics, University of Oviedo, Oviedo, Spain
bazjuan@uniovi.es, montes@uniovi.es
[2] School of Information Technology, Deakin University, Geelong, Australia
gleb@deakin.edu.au
[3] Department of Computer Science, University of Oviedo, Gijón, Spain
sirene@uniovi.es

Abstract. Aggregations of random variables have appeared as a way to model aggregation processes in which the input data is seen as observations of random variables. However, its definition includes many other scenarios that have not yet been studied in detail. This paper is devoted to define and study several families of aggregations of random variables that go beyond the aggregation of random inputs. Different characterizations, the relation between them and some illustrative examples are provided.

Keywords: Aggregation · Random variables · Probability theory

1 Introduction

Aggregation theory is focused on functions that summarize the information of several elements into a unique element. This type of functions are widely used in data analysis (see [7, 13]), in which working with random variables is the common approach made by Statistics. In [2], the concept of aggregation of random variables is defined as a way to study the aggregation when dealing with random inputs. Notice that this approach differs from others more usually considered in the literature, such as the aggregation of probability distributions, possibilities distributions or aggregation in Dempster-Shafer theory. We refer the reader to [5] for a survey in this regard.

However, the definition of aggregation of random variables is not too restrictive, and allows many type of functions to be considered as aggregations of random variables. In this paper, we define some families of aggregations of random variables that permit, for instance, to apply a real function after identifying the distribution of the inputs, to consider a change of the dependence between

* The authors have been supported by Programa Severo Ochoa of Principality of Asturias (BP21042), the Ministry of Science and Innovation (PID2022-139886NB-I00) and the Australian Research Council Discovery Project (DP210100227).

the inputs and the output and to work with randomly behaved parameters. Their characterization and the relation between them are studied in detail.

The rest of the paper is organized as follows. In Section 2, we introduce some basic notions about probability theory and aggregation of random variables. In Section 3 we introduce the class of conditionally determined. Then, aggregations with the same distribution are studied in Section 4. Section 5 is devoted to define randomly induced aggregations of random variables. Finally, in Section 6 we study the relation between the introduced families and the conclusions are discussed in Section 7.

2 Preliminaries

2.1 Probability notions

Along the paper, we will consider a fixed probability space (Ω, Σ, P), where Ω is a set, Σ a σ-algebra of measurable subsets of Ω and P a probability measure (see, for instance, [14]). We will suppose the probability space to be the right continuous adapted probability space associated with a hyperfinite adapted probability space. Informally, they are probability spaces much bigger than the usual unit interval with the Lebesgue measure that have the properties we need for some of the constructions we are going to address. We refer the reader to [9] and [10] for a proper introduction of such spaces. Before explaining more in detail the properties of this type of probability space, let us introduce the usual stochastic order.

Definition 1. *[15] Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors of dimension n. Then, if for any increasing function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $E[\phi(\boldsymbol{X})]$ and $E[\phi(\boldsymbol{Y})]$ exist it holds $E[\phi(\boldsymbol{X})] \leq_{st} E[\phi(\boldsymbol{Y})]$, it is said that $\boldsymbol{X}$ is smaller than or equal to $\boldsymbol{Y}$ with respect to the usual stochastic order, denoted as $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$.*

For random variables, the usual stochastic order is equivalent to the pointwise comparison of the distribution functions [15]. If $\boldsymbol{X} =_{st} \boldsymbol{Y}$, it is said that the random vectors have the same distribution. A notion related with the usual stochastic order is to be almost surely smaller than or equal to, denoted by $\boldsymbol{X} \leq_{a.s.} \boldsymbol{Y}$ and defined as $P(\boldsymbol{X} \leq \boldsymbol{Y}) = 1$. It is known, see [15], that $\boldsymbol{X} \leq_{a.s.} \boldsymbol{Y} \implies \boldsymbol{X} \leq_{st} \boldsymbol{Y}$. Another sufficient condition for having the usual stochastic order can be stated by using conditional distributions.

Proposition 1. *[15] Let $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be three random vectors. If $[\boldsymbol{X} \mid \boldsymbol{Z} = \boldsymbol{z}] \leq_{st} [\boldsymbol{Y} \mid \boldsymbol{Z} = \boldsymbol{z}]$ for any $\boldsymbol{z}$ in the support of $\boldsymbol{Z}$, then $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$.*

Returning to hyperfinite adapted probability spaces, one of the most important properties is that two random variables with the same distribution are linked by a measure preserving function. We recall that a measurable function between two measurable spaces is defined as a function which has measurable preimages of measurable sets (see [4]).

Definition 2. *[10] Let (Ω, Σ, P) be a probability space and let $\phi : \Omega \to \Omega$ be a measurable function. Then, it is said that ϕ is a measure preserving function if $P(\phi^{-1}(B)) = P(B)$ for any $B \in \Sigma$.*

Theorem 1. *[9] Let (Ω, Σ, P) be the right continuous adapted probability space associated with a hyperfinite adapted probability space. Then, given two random variables X and Y such that $X =_{st} Y$, there exists a bijective measure preserving function $\phi : \Omega \to \Omega$ such that $X =_{a.s} Y \circ \phi$.*

The function ϕ in the latter result does not have an easy expression. We refer the reader to page 134, Proposition 9.2 in [8] for a constructive proof. In addition, in [9] it is proved that any right continuous adapted probability space associated with a hyperfinite adapted probability space fulfills the saturation property, defined as follows.

Definition 3. *[11] A probability space (Ω, Σ, P) is said to have the saturation property if, given a pair of Polish spaces (complete metrizable topological spaces) T_1 and T_2, for any probability measure μ in $T_1 \times T_2$ and a random element $X : \Omega \to T_1$ such that its distribution coincides with the marginal distribution of μ in T_1, there exists a random element $Y : \Omega \to T_2$ such that (X, Y) has distribution function μ.*

For the scope of this paper, it is enough to consider as Polish spaces the Euclidean Spaces $\mathbb{R}^n$ with the usual topology. We end this section by introducing the concept of σ-algebra generated by a random vector.

Definition 4. *[4] Let $\boldsymbol{X}$ a random vector defined in a probability space (Ω, Σ, P). Then, the σ-algebra generated by $\boldsymbol{X}$, denoted as $\sigma(\boldsymbol{X})$, is the smallest σ-algebra such that $\boldsymbol{X}$ is measurable in $(\Omega, \sigma(\boldsymbol{X}))$.*

2.2 Aggregations of random variables

Given a non-empty real interval I, a function $\hat{A} : I^n \to I$ is an aggregation function if it is increasing and fulfills $\inf I = \inf_{\boldsymbol{x} \in I^n} \hat{A}(\boldsymbol{x})$ and $\sup I = \sup_{\boldsymbol{x} \in I^n} \hat{A}(\boldsymbol{x})$. Aggregations of random variables are functions that map, given a real interval I, a random vector with support I^n to a random variable with support I. The properties of monotonicity and boundary conditions are redefined in terms of stochastic orders. Before stating the definition, let us introduce the following notation.

$$L_I^n(\Omega) = \{\boldsymbol{X} : \Omega \to I^n \mid \boldsymbol{X} \text{ is measurable}\}.$$

If $n = 1$, we will denote $L_I^1(\Omega)$ just as $L_I(\Omega)$. In addition, we will just use the notation L_I^n, assuming that the probability space is fixed and supposed to be the right continuous adapted probability space associated with a hyperfinite adapted probability space.

Definition 5. *[2] Let (Ω, Σ, P) be a probability space and let I be a real non empty interval. An aggregation function of random variables (with respect to $\leq_{st}$) is a function $A : L_I^n \to L_I$ which satisfies:*

- *For any* $\boldsymbol{X}, \boldsymbol{Y} \in L_I^n(\Omega)$ *such that* $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$, $A(\boldsymbol{X}) \leq_{st} A(\boldsymbol{Y})$.
- *For any* $X \in L_I$, *there exist* $\boldsymbol{X}_1, \boldsymbol{X}_2 \in L_I^n$ *such that* $A(\boldsymbol{X}_1) \leq_{st} X \leq_{st} A(\boldsymbol{X}_2)$.

The second property in the latter definition is known as the boundary conditions. We want to remark that, in the original reference [2], these conditions are defined in a different but equivalent manner.

Since this paper is devoted to the classification of aggregations of random variables, we present in the next definition three already known families that have been proven to be disjoint when the interval do not consist on just one point (see [3]), although their union is not the whole set of aggregations of random variables.

Definition 6. *[3] Let* $A : L_I^n \to L_I$ *be an aggregation of random variables. Then,*

- *If* $A(\boldsymbol{X}) = \hat{A} \circ \boldsymbol{X}$ *for any* $\boldsymbol{X} \in L_I^n$ *with* $\hat{A} : I^n \to I$ *being an aggregation function, A is said to be induced (by* $\hat{A}$*).*
- *If* $A(\boldsymbol{X})$ *has degenerate distribution for any* $\boldsymbol{X} \in L_I^n$, *A is said to be degenerate.*
- *If there exists* $\boldsymbol{X} \in L_I^n$ *and* $\boldsymbol{x}$ *in the support of* $\boldsymbol{X}$ *such that* $[A(\boldsymbol{X}) \mid \boldsymbol{X} = \boldsymbol{x}]$ *is not degenerate, A is said to be random.*

Induced aggregations of random variables are just usual aggregation functions with random inputs. Degenerate aggregations of random variables appear when we aggregate location parameters of the associated random vectors and the random ones represent that there is a random behavior in the aggregation process, not only in the inputs.

3 Conditional determination

Analyzing the definition of random aggregation of random variables, we can see that it is based on a negation of a property. In particular, we need the existence of a random vector $\boldsymbol{X} \in L_I^n$ and a value $\boldsymbol{x} \in I^n$ such that $[A(\boldsymbol{X})| \, \boldsymbol{X} = \boldsymbol{x}]$ is well defined and not degenerate. Let us consider the class of aggregations of random variables that do fulfill this property.

Definition 7. *Let* $A : L_I^n \to L_I$ *be an aggregation of random variables. It is said that A is conditionally determined if for any* $\boldsymbol{X} \in L_I^n$ *and* $\boldsymbol{x} \in I^n$ *for which* $[A(\boldsymbol{X}) \mid \boldsymbol{X} = \boldsymbol{x}]$ *is well defined,* $[A(\boldsymbol{X}) \mid \boldsymbol{X} = \boldsymbol{x}]$ *has degenerate distribution.*

Being conditionally determined can be interpreted as having the value of the output of the aggregation totally determined when knowing the input random and its value. We want to remark that, unlike induced aggregations, the value could change depending on the input random vector. Trivially, any aggregation of random variables is conditionally determined or random. Let us introduce an example in this regard.

Example 1. Given a fixed value $z \in I$, let $A : I^n \to I$ be the aggregation of random variables given by $A(\boldsymbol{X}) = \min(\boldsymbol{X})$ if $\boldsymbol{X} \leq_{st} x\mathbf{1}$, $A(\boldsymbol{X}) = \max(\boldsymbol{X})$ if $\boldsymbol{X} \leq_{st} x\mathbf{1}$ and $A(\boldsymbol{X}) = x\mathbf{1}$ otherwise. A is conditionally determined, but is not induced.

In the following result, we give two alternatives characterizations of this type of aggregations of random variables.

Theorem 2. *Let $A : L_I^n \to L_I$ be an aggregation of random variables. The following properties are equivalent.*

(1) A is conditionally determined.
(2) For any $\boldsymbol{X}$, $A(\boldsymbol{X})$ is $\sigma(\boldsymbol{X})$-measurable.
(3) There exists a family of functions $(G_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ such that $A(\boldsymbol{X}) =_{a.s.} G_{\boldsymbol{X}} \circ \boldsymbol{X}$ for any $\boldsymbol{X} \in L_I^n$.

Proof. Suppose that A fulfills (1). For any $\boldsymbol{X} \in L_I^n$, let $C_{\boldsymbol{X}} = \{\boldsymbol{x} \in I^n \mid [A(\boldsymbol{X}) \mid \boldsymbol{X} = \boldsymbol{x}]$ is well-defined$\}$. Define $G_{\boldsymbol{X}} : I^n \to I$ as the function such that $G_{\boldsymbol{X}}(\boldsymbol{x}) = \lambda$ with λ the value that $[A(\boldsymbol{X}) \mid \boldsymbol{X} = \boldsymbol{x}]$ takes with probability 1 if $\boldsymbol{x} \in C_{\boldsymbol{X}}$ and $G_{\boldsymbol{X}}(\boldsymbol{x}) = 0$ otherwise. Since $P(\boldsymbol{X} \in C) = 1$, it is concluded that $A(\boldsymbol{X}) =_{a.s.} G_{\boldsymbol{X}} \circ \boldsymbol{X}$ for any $\boldsymbol{X} \in L_I^n$. Then (3) holds.

Suppose that (3) holds. Then, for any measurable set B of $\mathbb{R}$, one has $(G_{\boldsymbol{X}}(\boldsymbol{X}))^{-1}(B) = \boldsymbol{X}^{-1}(G_{\boldsymbol{X}}^{-1}(B)) \in \sigma(\boldsymbol{X})$. Then, (2) holds.

Suppose that (2) holds. Then, given $\boldsymbol{X} = \boldsymbol{x}$, one has that $\boldsymbol{X}^{-1}(\boldsymbol{x})$ is a measurable set that does not contains any other measurable set (in $\sigma(\boldsymbol{X})$). Then, $A(\boldsymbol{X})$ should take an unique value on $\boldsymbol{X}^{-1}(\boldsymbol{x})$. Then, A is conditionally determined and (1) holds. □

The third point in last result gives us a comprehensible characterization of conditional determination. We first identify the random vector $\boldsymbol{X}$ we are aggregating and select a particular function $G_{\boldsymbol{X}}$ and then we apply it to $\boldsymbol{X}$ to obtain the aggregated random variable. Of course, not all choices of $(G_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ are suitable to define an aggregation of random variables. Measurability, monotonicity and the boundary conditions should be guaranteed.

This property appears naturally in some areas of statistics. For mean estimation, the function that is applied to the random sample varies depending on the distribution. For instance, the best estimator for Gaussian distributions is the arithmetic mean and for uniform distributions is the average between the maximum and the minimum [14].

Notice that induced and degenerate aggregations of random variables are contained in the conditionally determined ones. In particular, if $G_{\boldsymbol{X}} = \hat{A}$ with $\hat{A}$ an usual aggregation function for all $\boldsymbol{X} \in L_I^n$, then A is induced. Similarly, if for any $\boldsymbol{X} \in L_I^n$ one has that $G_{\boldsymbol{X}}$ only takes one value, A is degenerate.

4 Equality in distribution

In the previous section, we gave an understandable characterization of conditionally determined aggregations of random variables. However, it remains to study more in detail the random ones, which can be seen as aggregations with a random behavior in the aggregation process.

If we look at Definition 5, the monotonicity focuses on the distribution of $\boldsymbol{X}$ and $A(\boldsymbol{X})$, but not on the dependence between them. In particular, the monotonicity implies $\boldsymbol{X} =_{st} \boldsymbol{Y} \implies A(\boldsymbol{X}) =_{st} A(\boldsymbol{Y})$, but it does not imply neither $\boldsymbol{X} =_{st} \boldsymbol{Y} \implies (\boldsymbol{X}, A(\boldsymbol{X})) =_{st} (\boldsymbol{Y}, A(\boldsymbol{Y}))$ nor $\boldsymbol{X} =_{a.s.} \boldsymbol{Y} \implies A(\boldsymbol{X}) =_{a.s.} A(\boldsymbol{Y})$. This allows some particular type of aggregation of random variables to be defined. Let us start with a simple example.

Example 2. Let $A : L_I^n \to L_I$ be an induced aggregation. Consider U a uniform random variable. For any $\boldsymbol{X} \in L_I^n$, define $B(\boldsymbol{X}) = F^{-1}_{A(\boldsymbol{X})}(U)$, where $F^{-1}_{A(\boldsymbol{X})}$ is the quantile function of $A(\boldsymbol{X})$. Trivially, $B(\boldsymbol{X}) =_{st} A(\boldsymbol{X})$. Thus $B : L_I^n \to L_I$ is an aggregation of random variables. However, B is not induced. Moreover, is random, because the value of the aggregated random variable depends on U.

In the latter example, we construct a random aggregation of random variables in which the outputs have the same distribution as another one, but the dependence between them and the inputs is different. In this direction, given $A, B : L_I^n \to L_I$ two aggregations of random variables, if $A(\boldsymbol{X}) =_{st} B(\boldsymbol{X})$ for any $\boldsymbol{X} \in L_I^n$, we will say that A and B have the same distribution. In the following result, we study the extension of the family of conditionally determined aggregations of random variables by considering the aggregations that have the same distribution. We recall that the set of mass points of a random vector are the values that the random vector takes with probability greater than 0.

Theorem 3. *Let $A : L_I^n \to L_I$ be an aggregation of random variables. The following properties are equivalent.*

(1) A has the same distribution as a conditionally determined aggregation of random variables.
(2) There exists a family of measure preserving transformations $(\phi_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ and random vectors $(\boldsymbol{Z}_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ such that $A(\boldsymbol{X}) \circ \phi_{\boldsymbol{X}} =_{a.s.} \boldsymbol{Z}_{\boldsymbol{X}}$ and $\boldsymbol{Z}_{\boldsymbol{X}}$ is $(\sigma(\boldsymbol{X}))$-measurable for any $\boldsymbol{X} \in L_I^n$.
(3) For any $\boldsymbol{X} \in L_I^n$, if $S_{\boldsymbol{X}}$ denotes the set of probability of mass points of $\boldsymbol{X}$ and $S_{A(\boldsymbol{X})}$ denotes the probability of mass points of $A(\boldsymbol{X})$, there exists $L : S_{\boldsymbol{X}} \to S_{A(\boldsymbol{X})}$ such that:

$$\sum_{\boldsymbol{x} \in L^{-1}(x)} P(\boldsymbol{X} = \boldsymbol{x}) \leq P(A(\boldsymbol{X}) = x)$$

for any $x \in S_{A(\boldsymbol{X})}$

Proof. Suppose that (1) holds. Then, there exists a conditionally determined aggregation of random variables $B : L_I^n \to L_I$ such that $B(\boldsymbol{X}) =_{st} A(\boldsymbol{X})$ for any $\boldsymbol{X} \in L_I^n$. Then, (2) holds by using Theorem 1 and (2) in Theorem 2.

Suppose that (2) holds. For any $\boldsymbol{X} \in L_I^n$, since $\phi_{\boldsymbol{X}}$ is a measure preserving transformation, $A(\boldsymbol{X}) =_{st} A(\boldsymbol{X}) \circ \phi_{\boldsymbol{X}}$. Applying (2) on Theorem 2, $A(\boldsymbol{X}) \circ \phi_{\boldsymbol{X}}$ is conditionally determined, thus (1) holds.

Suppose that (1) holds. Applying (3) in Theorem 2, for any $\boldsymbol{X} \in L_I^n$, there exist a function $G_{\boldsymbol{X}}$ such that $G_{\boldsymbol{X}}(\boldsymbol{X}) =_{st} A(\boldsymbol{X})$. Define $L : S_{\boldsymbol{X}} \to S_{A(\boldsymbol{X})}$ as $L(\boldsymbol{x}) = G_{\boldsymbol{X}}(\boldsymbol{x})$. Then, $P(A(\boldsymbol{X}) = x) = P(\boldsymbol{X} \in L^{-1}(x)) \geq \sum_{\boldsymbol{x} \in L^{-1}(x)} P(\boldsymbol{X} = \boldsymbol{x})$ for any $x \in S_{A(\boldsymbol{X})}$. It is concluded that (3) holds.

Suppose that (3) holds. Decompose the distribution function of $\boldsymbol{X}$ as $F_{\boldsymbol{X}} = \lambda F_{\boldsymbol{X}}^d + (1-\lambda) F_{\boldsymbol{X}}^c$, with F_d being the distribution function of a discrete random vector, F_c being the distribution function of a continuous random vector and $\lambda = \sum_{\boldsymbol{x} \in S_{\boldsymbol{X}}} P(\boldsymbol{X} = \boldsymbol{x})$. Similarly, decompose the distribution function of $A(\boldsymbol{X})$ as $F_{A(\boldsymbol{X})} = \mu F_{A(\boldsymbol{X})}^d + (1-\mu) F_{A(\boldsymbol{X})}^c$ with $\mu = \sum_{x \in S_{\boldsymbol{X}}} P(A(\boldsymbol{X}) = x)$. Given a random vector $\boldsymbol{Y}$ with distribution function $F_{\boldsymbol{X}}^d$, consider the transformation $L(\boldsymbol{Y})$ with L given by (3). In addition, given a random vector $\boldsymbol{Z}$ with distribution function $F_{\boldsymbol{X}}^c$, the first component Z_1 has a continuous distribution function F_1, and $F_1(Z_1)$ is a uniform random variable. By the inequality given in (3), one has that $\lambda \leq \mu$, that is, the discrete part of $\boldsymbol{X}$ is smaller than the discrete part of $A(\boldsymbol{X})$. Therefore, $F_{A(\boldsymbol{X})} = \lambda F_{L(\boldsymbol{Y})} + (\mu - \lambda)\hat{F} + (1-\mu) F_{A(\boldsymbol{X})}^c$, where (in the case of $\mu - \lambda \neq 0$) if W has distribution $\hat{F}$ one has that $P(W = x) = \frac{1}{\mu - \lambda} \left(P(A(\boldsymbol{X}) = x) - \sum_{\boldsymbol{x} \in L^{-1}(x)} P(\boldsymbol{X} = \boldsymbol{x}) \right) \geq 0$. If $\lambda \neq 1$, denote $F_0 = \frac{1}{1-\lambda}(\mu - \lambda)\hat{F} + (1-\mu) F_{A(\boldsymbol{X})}^c$ and consider the corresponding quantile function F_0^{-1}. Now, consider the function $G_{\boldsymbol{X}} : \mathbb{R}^n \to \mathbb{R}$ defined as $G_{\boldsymbol{X}}(\boldsymbol{x}) = L(\boldsymbol{x})$ if $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and $G_{\boldsymbol{X}}(\boldsymbol{x}) = F_0^{-1}(F_1(x_1))$ in other case. By construction, it is clear that $G_{\boldsymbol{X}}$ has distribution function $F_{G_{\boldsymbol{X}}(\boldsymbol{X})} = \lambda F_{L(\boldsymbol{Y})} + (1-\lambda) F_0 = \lambda F_{L(\boldsymbol{Y})} + (\mu - \lambda)\hat{F} + (1-\mu) F_{A(\boldsymbol{X})}^c = F_{A(\boldsymbol{X})}$. If $\lambda = 1$, then F_0 can be any function. It is concluded that, for any $\boldsymbol{X} \in L_I^n$, there exists a family of functions $(G_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ such that $G_{\boldsymbol{X}}(\boldsymbol{X}) =_{st} A(\boldsymbol{X})$ and, applying (3) in Theorem 2, that (1) holds. □

The change of the dependence between the inputs and the output can be expressed in terms of the measure preserving transformations that appear in the second characterization of the latter result. These transformations play a very important role in the study of stationary time series, in particular in ergodic theory (see [12]).

Informally, the last characterization says that if we can find functions that fit the probability mass points of the inputs in the probability mass points of the outputs, then we have an aggregation of random variables that has the same distribution as a conditionally determined one. The most simple example of a scenario in which this not happens is when a degenerate random vector has associated an aggregated random variable that is continuous.

5 Random parameters

As explained before, not all aggregations of random variables have the same distribution as a conditionally determined one. One of the cases in which this happens is when we have random parameters in the aggregation. For instance, in a weighted mean with continuous random parameters, the aggregated random variables can be continuous even in the case of a degenerate input.

Random parameters appear in many real life situations. One of the most important scenarios is the one in which the parameters of a family of aggregation functions are fitted using data, see, for instance, [13]. If we consider the training data as realizations of random variables, then also the parameters have a random behavior. In addition, there are scenarios in which the theoretical parameters are considered to be random, see [1].

In the construction of aggregations of random variables with random parameters, one may think to model the random parameters by considering a random vector. However, this election is not adequate in terms of monotonicity. Let us illustrate this problem with a simple example.

Example 3. Let U_1, U_2 and U_3 be three standard independent uniform random variables. Consider the function $A : L_I^2 \to L_I$ such that $A(X_1, X_2) = U_1 X_1 + (1 - U_1) X_2$. It can be seen as a weighted mean with countermonotone (with perfect negative dependence) and uniform weights. However, the monotonicity cannot be fulfilled because $(U_1, U_2) =_{st} (U_2, U_3)$ but $A(U_1, U_2) = U_1^2 + (1 - U_1)U_2 \neq_{st} U_1 U_2 + (1 - U_1) U_3 = A(U_2, U_3)$.

As illustrated in the last example, the main problem of random parameters fixed as a random vector is that one can have inputs with the same distribution but a different dependence with the random parameters, resulting in a different output distribution, which breaks the monotonicity.

A solution for that is to fix the distribution of the random parameters and their dependence with the inputs and construct, for each of the cases, a random vector fulfilling these properties. Some sufficient conditions are given in the next result.

Theorem 4. *Let I be a real interval and $\hat{A} : I^n \times \mathbb{R}^d \to I$ be a measurable function and let $(\boldsymbol{\lambda}_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ be a family of random vectors such that:*

- *For any $\boldsymbol{z} \in \mathbb{R}^d$, the function $\hat{B}_{\boldsymbol{z}} : I^n \to I$ defined as $\hat{B}_{\boldsymbol{z}}(x_1, \dots, x_n) = \hat{A}(x_1, \dots, x_n, z_1, \dots, z_d)$ is an aggregation function.*
- *If I does not have a lower [upper] bound, for any $x \in I$ there exists $\boldsymbol{x} \in I^n$ such that $\hat{A}(\boldsymbol{x}, \boldsymbol{z}) < [>] x$ for any $\boldsymbol{z} \in \mathbb{R}^n$.*
- *$\boldsymbol{\lambda}_{\boldsymbol{X}}$ has the same distribution for any $\boldsymbol{X} \in L_I^n$*
- *$\boldsymbol{X} \leq_{st} \boldsymbol{Y} \implies [\boldsymbol{X} \mid \boldsymbol{\lambda}_{\boldsymbol{X}} = \boldsymbol{z}] \leq_{st} [\boldsymbol{Y} \mid \boldsymbol{\lambda}_{\boldsymbol{Y}} = \boldsymbol{z}]$ for any $\boldsymbol{X}, \boldsymbol{Y} \in L_I^n$.*

Then, the function $A : L_I^n \to L_I$ defined as $A(\boldsymbol{X}) = \hat{A}(\boldsymbol{X}, \boldsymbol{\lambda}_{\boldsymbol{X}})$ is an aggregation of random variables.

Proof. Noticing that since $\hat{A}$ is measurable and its image is I, it is clear that $A : L_I^n \to L_I$ is well-defined. For the monotonicity, use that $\hat{B}_{\boldsymbol{z}}$ is increasing for any $\boldsymbol{z} \in \mathbb{R}^n$, $\hat{B}_{\boldsymbol{z}}([\boldsymbol{X} \mid \boldsymbol{\lambda}_{\boldsymbol{X}} = \boldsymbol{z}]) \leq_{st} \hat{B}_{\boldsymbol{z}}([\boldsymbol{Y} \mid \boldsymbol{\lambda}_{\boldsymbol{Y}} = \boldsymbol{z}])$ if well-defined. Finally, since $\lambda_{\boldsymbol{X}} =_{st} \lambda_{\boldsymbol{Y}}$, applying Proposition 1, one has that $A(\boldsymbol{X}) = \hat{A}(\boldsymbol{X}, \boldsymbol{\lambda}_{\boldsymbol{X}}) = \hat{B}_{\boldsymbol{\lambda}_{\boldsymbol{X}}}(\boldsymbol{X}) \leq_{st} \hat{B}_{\boldsymbol{\lambda}_{\boldsymbol{Y}}}(\boldsymbol{Y}) = \hat{A}(\boldsymbol{Y}, \boldsymbol{\lambda}_{\boldsymbol{Y}}) = A(\boldsymbol{Y})$.

For the boundary conditions, suppose that I has a lower bound a. If $\boldsymbol{X} =_{st} (a, \ldots, a)$, by the boundary conditions of $\hat{A}_{\boldsymbol{z}}$ one has that $\hat{A}_{\boldsymbol{z}}(\boldsymbol{X}) =_{st} a$ for any $\boldsymbol{z} \in \mathbb{R}^d$, thus $A(\boldsymbol{X}) =_{st} a$.

If I has not a lower bound, consider $X \in L_I$. For each $k \in I \cap \mathbb{Z}$, consider $\boldsymbol{x}_k \in I^n$ such that $\hat{A}(\boldsymbol{x}, \boldsymbol{z}) < x$ for any $\boldsymbol{z} \in \mathbb{R}^n$, that exists by hyphotesis. Then, define the function $h : I \to I^n$ as $h(x) = \boldsymbol{x}_k$ if $x \in [k, k+1) \cap I$. Notice that since h is measurable and its image is I^n, one has that $h(X) \in L_I^n$. By construction, $A(h(X)) <_{a.s.} X$ and it is concluded that for any $X \in L_I$ there exists $\boldsymbol{X} \in L_I^n$ such that $A(\boldsymbol{X}) \leq_{st} X$. For the upper bound, the proof is the same. □

The conditions over $(\boldsymbol{\lambda_X}, \boldsymbol{X} \in L_I^n)$ might seem very strong, but a simple case in which they are hold is when $\boldsymbol{X}$ and $\boldsymbol{\lambda_X}$ are independent for any $\boldsymbol{X} \in L_I^n$. Notice that these type of structures can be defined in our probability space since it fulfills the saturation property introduced in Definition 3. More involved cases maybe can be found by fixing a vector copula (see [6]) between $\boldsymbol{X}$ and $\boldsymbol{\lambda_X}$. For the second condition of the theorem, it is enough to consider, for instance, the aggregations $\hat{B}_{\boldsymbol{z}}$ to be internal, i.e. between the maximum and the minimum.

Random induced aggregations of random variables includes the induced ones and are not contained in the ones that have the same distribution as a conditionally determined. Of course, random parameters can be combined with the change of dependence considered in Section 4 or the consideration of a family of functions in Section 3.

For the first one, we can consider the family of aggregations of random variables that have the same distribution as a randomly induced aggregation of random variable. This family is bigger, and it does not contain the set of conditionally determined aggregation functions, since Example 2 is not included.

In the second case, if we consider the aggregations of random variables such that exist a family of random vectors $(\boldsymbol{\lambda_X}, \boldsymbol{X} \in L_I^n)$ and $(G_{\boldsymbol{X}}, \boldsymbol{X} \in L_I^n)$ such that $A(\boldsymbol{X}) =_{a.s.} G_{\boldsymbol{X}}(\boldsymbol{X}, \boldsymbol{\lambda_X})$, we obtain the set of all aggregations of random variables (for the considered interval I). In particular, it suffices to consider $\boldsymbol{\lambda_X}$ to be an univariate random variable that is almost surely equal to any possible value of $A(\boldsymbol{X})$ and $G_{\boldsymbol{X}}(\boldsymbol{X}, \boldsymbol{\lambda_X}) = \boldsymbol{\lambda_X}$ for any $\boldsymbol{X} \in L_I^n$.

6 Relation between the families of aggregations of random variables

As already mentioned before, induced and degenerate aggregations of random variables are two disjoint families that are contained in the conditionally determined ones. In the next result, the relation between other families is studied.

Proposition 2. *Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then,*

(1) A is almost surely equal to an induced aggregation if and only if is conditionally determined and is almost surely equal to a randomly induced aggregation.
(2) A has the same distribution as an induced aggregation of random variables if and only if it has the same distribution as a conditionally determined and a randomly induced aggregations of random variables.

(3) If A has the same distribution as a conditionally determined aggregation of random variables and is randomly induced, then it is induced.
(4) If A has the same distribution as a degenerate aggregation of random variables, then it is degenerate.

Proof. For (1), proving that being induced implies being conditionally determined and randomly induced is straightforward. If A is almost surely equal to a randomly induced aggregation, then $A(\boldsymbol{X}) =_{a.s.} \hat{A}(\boldsymbol{X}, \boldsymbol{\lambda_X})$ for any $\boldsymbol{X} \in L_I^n$ with $\hat{A} : I^n \times \mathbb{R}^d \to I$ and $(\boldsymbol{\lambda_X}, \boldsymbol{X} \in L_I^n)$ as in Theorem 4. If A is also conditionally determined, applying (3) in Theorem 2 we also have $A(\boldsymbol{X}) =_{a.s.} G_{\boldsymbol{X}}(\boldsymbol{X})$ for any $\boldsymbol{X} \in L_I^n$. Then, $\hat{A}(\boldsymbol{X}, \boldsymbol{\lambda_X}) =_{a.s} G_{\boldsymbol{X}}(\boldsymbol{X})$. The expectation $E[\hat{A}(\boldsymbol{x}, \boldsymbol{\lambda_X})]$, since equals $G_{\boldsymbol{X}}(\boldsymbol{X})$, always exists for any $\boldsymbol{x} \in I^n$ and $\boldsymbol{X} \in L_I^n$. In addition, since $\boldsymbol{\lambda_X}$ has the same distribution for all $\boldsymbol{X} \in L_I^n$, $E[\hat{A}(\boldsymbol{x}, \boldsymbol{\lambda_X})]$ takes always the same value for a fixed $\boldsymbol{x} \in I^n$ and any $\boldsymbol{X} \in L_I^n$. Then, define $\hat{B} : I^n \to I$ as $\hat{B}(\boldsymbol{x}) = E[\hat{A}(\boldsymbol{x}, \boldsymbol{\lambda_X})]$. Since $\hat{A}$ is increasing in the first n components, $E[\hat{A}(\boldsymbol{x}_1, \boldsymbol{\lambda_X})] \leq E[\hat{A}(\boldsymbol{x}_2, \boldsymbol{\lambda_X})]$ if $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$. If I has a lower bound a, then it is clear that $E[\hat{A}(a\mathbf{1}, \boldsymbol{\lambda_X})] = a$. If I does not have a lower bound, for any $x \in I$ there exists $\boldsymbol{x} \in I^n$ such that $\hat{A}(\boldsymbol{x}, \boldsymbol{z}) < x$ for any $\boldsymbol{z} \in \mathbb{R}^n$. Then, $E[\hat{A}(\boldsymbol{x}_1, \boldsymbol{\lambda_X})] < x$. Proceeding similarly for the upper bound, we have that $\hat{B}$ is increasing and fulfils the boundary conditions, thus is an aggregation function. It is concluded that $A(\boldsymbol{X}) =_{a.s.} G_{\boldsymbol{X}}(\boldsymbol{X}) =_{a.s} \hat{B} \circ \boldsymbol{X}$ with $\hat{B}$ being an aggregation function.

The proof of (2) is equivalent as the latter one but replacing $=_{a.s.}$ by $=_{st}$.

Let us now prove (3). For any $\boldsymbol{x} \in I^n$, consider the random vector $\boldsymbol{X}$ such that $P(\boldsymbol{X} = \boldsymbol{x}) = 1$, then since A has the same distribution as a conditionally determined aggregation of random variables, $A(\boldsymbol{X})$ should be degenerate. But, since it is randomly induced, one has that $A(\boldsymbol{X}) = \hat{A}(\boldsymbol{X}, \lambda_{\boldsymbol{X}})$. Recall that $\boldsymbol{\lambda_X}$ has the same distribution for any $\boldsymbol{X} \in L_I^n$. If the distribution of $\lambda_{\boldsymbol{X}}$ is not degenerate, then there exists a function $\hat{B} : I^n \to I$ such that $\hat{B}(\boldsymbol{x}) = \hat{A}(\boldsymbol{x}, \boldsymbol{z})$ for any $\boldsymbol{z} \in \mathbb{R}^d$. If $\lambda_{\boldsymbol{X}}$ is degenerate with $P(\boldsymbol{\lambda_X} = \boldsymbol{z}) = 1$, then define $\hat{B} : I^n \to I$ as $\hat{B}(\boldsymbol{x}) = \hat{A}(\boldsymbol{x}, \boldsymbol{z})$. In both cases, it is clear that $A(\boldsymbol{X}) = \hat{B}(\boldsymbol{X})$ and A is induced.

Finally, for (4) consider B a degenerate aggregation of random variables such that $A(\boldsymbol{X}) =_{st} B(\boldsymbol{X})$ for any $\boldsymbol{X} \in L_I^n$. If for any $\boldsymbol{X} \in L_I^n$ there exists $x \in I$ such that $P(B(\boldsymbol{X}) = x) = 1$, then it is straightforward that $P(A(\boldsymbol{X}) = x) = 1$ and, therefore, A is degenerate. □

It remains to see if there exists a conditionally determined aggregation such that it has the same distribution as an induced one but it is not induced. We provide an example in this regard by using countermonotone (with perfect negative dependence) random variables.

Example 4. Let $A : L_I^n \to L_I$ be an aggregation of random variables defined as $A(\boldsymbol{X}) = \max(\boldsymbol{X})$ for any $\boldsymbol{X}$ such that $\max(\boldsymbol{X})$ is not continuous and $A(\boldsymbol{X}) = Y$ with Y being a random variable that has the same distribution as $\max(\boldsymbol{X})$ and such that Y and $\max(\boldsymbol{X})$ are countermonotone if $\max(\boldsymbol{X})$ is continuous. Then, A has the same distribution as the induced random variable $\max(\boldsymbol{X})$, is conditionally determined since $A(\boldsymbol{X})$ can be expressed as a function of $\max(\boldsymbol{X})$, that is a function of $\boldsymbol{X}$, but is not induced.

Now, with all the relations between the different families already studied, we can represent them as in Figure 1, in which each family is associated to a subset of the plane and in which the intersection of those sets represent the intersection of the families.

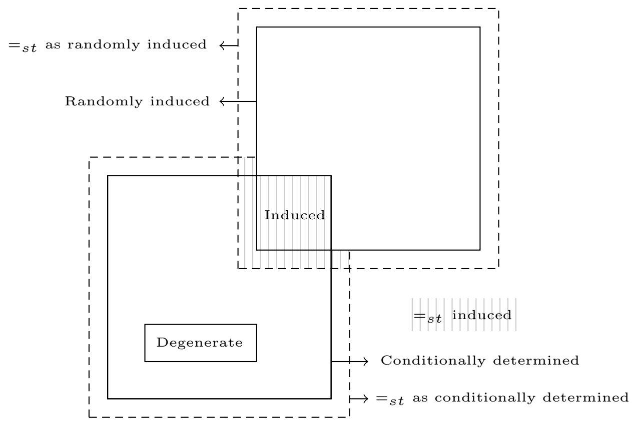

Fig. 1. Representation of families of aggregations of random variables.

In addition, the study is also useful to identify the different scenarios that aggregations of random variables can model. In particular, we have seen that, in addition to aggregations with random inputs, it is possible to identify the random vector we are aggregating in order to change the function we apply, to change the dependence between the inputs and the output and to consider random parameters. In Figure 2, the resulting aggregations of random variables that appear when we consider the different scenarios are represented.

7 Conclusions and future work

In this paper, several families of aggregations of random variables are defined, characterized and their relations are studied. These families can be seen as examples in which the concept of aggregation of random variables goes further than to simply apply usual aggregation functions to random inputs. Some examples, among others, include the application of a real function that changes depending on the distribution of the input, a procedure that is common in Statistics, and the incorporation of randomness in the aggregation process by the uncertainty of fitted parameters.

As an open question, we wonder if this classification still makes sense when considering the aggregation of other random structures such that stochastic processes, random elements in bounded lattices or random sets.

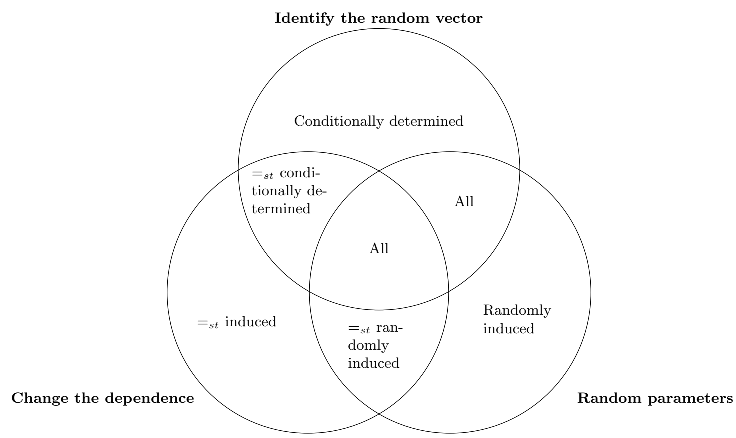

Fig. 2. Families of aggregations of random variables that appear when identifying the random vector, changing the dependence and/or considering random parameters.

References

1. Asimit, A.V., Hashorva, E., Kortschak, D.: Aggregation of randomly weighted large risks. IMA Journal of Management Mathematics **28**(3), 403–419 (2017)
2. Baz, J., Díaz, I., Montes, S.: The choice of an appropriate stochastic order to aggregate random variables. In: International Conference on Soft Methods in Probability and Statistics. pp. 40–47. Springer (2022)
3. Baz, J., Díaz, I., Montes, S.: Aggregation of random elements over bounded lattices. International Journal of Approximate Reasoning p. 109122 (2024)
4. Bogachev, V.I.: Measure theory. Springer Science & Business Media (2007)
5. Dubois, D., Liu, W., Ma, J., Prade, H.: The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. Information Fusion **32**, 12–39 (2016)
6. Fan, Y., Henry, M.: Vector copulas. Journal of Econometrics **234**(1), 128–150 (2023)
7. James, S.: An introduction to data analysis using aggregation functions in R. Springer (2016)
8. Keisler, H.J.: An infinitesimal approach to stochastic analysis, vol. 297. American Mathematical Soc. (1984)
9. Keisler, H.J.: Hyperfinite models of adapted probability logic. Annals of pure and applied logic **31**, 71–86 (1986)
10. Keisler, H.J.: Rich and saturated adapted spaces. Advances in Mathematics **128**(2), 242–288 (1997)
11. Keisler, H.J., Sun, Y.: Why saturated probability spaces are necessary. Advances in Mathematics **221**(5), 1584–1607 (2009)
12. Petersen, K.E.: Ergodic theory, vol. 2. Cambridge university press (1989)
13. Pollesch, N., Dale, V.H.: Applications of aggregation theory to sustainability assessment. Ecological Economics **114**, 117–127 (2015)
14. Rohatgi, V.K., Saleh, A.M.E.: An introduction to probability and statistics. John Wiley & Sons (2015)
15. Shaked, M., Shanthikumar, J.G.: Stochastic orders. Springer (2007)