Universidad de Oviedo

Programa de Doctorado en Matemática y Estadística

Agregación de estructuras aleatorias
Aggregation of random structures

Juan Baz González

2025

Universidad de Oviedo

Departamento de Estadística e Investigación Operativa y Didáctica
de la Matemática

Programa de Doctorado en Matemática y Estadística

Agregación de estructuras aleatorias
Aggregation of random structures

Realizada por **Juan Baz González**
bajo la dirección de Susana Montes Rodríguez
y Susana Irene Díaz Rodríguez

2025

Universidad de Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Agregación de estructuras aleatorias	Inglés: Aggregation of random structures

2.- Autor	
Nombre: Juan Baz González	
Programa de Doctorado: Programa de Doctorado en Matemáticas y Estadística	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN en español (máximo 4000 caracteres)

En el contexto de análisis de datos, las funciones de agregación se utilizan para fusionar la información de diferentes fuentes en un único valor. En esta dirección, es razonable considerar una naturaleza aleatoria en los valores agregados, ya que usualmente corresponden a diferentes observaciones de magnitudes dentro de una población. Siguiendo esta idea, este trabajo se centra en la definición y estudio de agregaciones de estructuras aleatorias.

En primer lugar, se introduce el concepto de agregación de variables aleatorias adaptando la monotonía y las condiciones de contorno haciendo uso de órdenes estocásticos. La elección del orden estocástico es crucial, siendo la mejor alternativa el orden estocástico usual. Se hace un estudio en profundidad de las principales familias de agregaciones de variables aleatorias y algunas propiedades clásicas de funciones de agregación para este nuevo concepto.

En segundo lugar, se trabaja con diferentes casos particulares de agregaciones de variables aleatorias que son de especial interés. En particular, se considera un operador de fusión lineal ordenado e inducido en el contexto de estimación de la media, se prueban resultados sobre el comportamiento asintótico de los pesos óptimos acumulados de L-estadísticos, se definen agregaciones de variables aleatorias que ordenan las variables de entrada comparando sus funciones de distribución, se estudia el comportamiento de capacidades aleatorias uniformes y se resuelve un problema de minimización que permite realizar un tipo específico de aproximaciones de agregaciones de variables aleatorias.

Posteriormente, se hace un estudio del comportamiento de algunas funciones de agregación y funciones asociadas cuando sus entradas son variables aleatorias haciendo uso de órdenes estocásticos. Primero se prueban resultados sobre la reducción de la variabilidad, respecto al orden estocástico convexo, que sucede al aplicar algunas funciones promedio. Se concluye también que, para muchas medidas de variabilidad, estas toman valores más grandes cuando el vector aleatorio asociado es más disperso o tiene una dependencia negativa más fuerte.

El concepto de agregación de variables aleatorias se extiende a otras estructuras aleatorias, dedicando más atención al caso de elementos aleatorios sobre conjuntos parcialmente ordenados y acotados, vectores aleatorios, procesos estocásticos y conjuntos aleatorios. Además de la construcción de dichos conceptos, se concluye que varias propiedades de vectores aleatorios o procesos estocásticos se preservan al aplicar la misma función de agregación a, respectivamente, cada una de las componentes o de índices asociados. Respecto a los conjuntos aleatorios, se define un orden estocástico de localización para los mismos.

Finalmente, algunos de los métodos desarrollados se aplican a problemas que involucran datos, ya sean reales o simulados. En particular, se estudian dos problemas de predicción con datos de temperatura y humedad, se estudia el comportamiento de contrastes de hipótesis relacionados con la uniformidad de capacidades aleatorias, el orden estocástico dispersivo, el orden estocástico convexo y el orden estocástico supermodular y se estudian mediante

simulación métodos de estimación para la media de variables aleatorias simétricas con dimensión variable.

RESUMEN en Inglés

In the context of data analysis, aggregation functions are used to fuse information from different sources into a unique value. In this direction, it is reasonable to consider a random behavior in the aggregated values, since usually they are related to different observations of quantities over a population. Following this idea, the work is focused on the definition and study of aggregations of random structures.

Firstly, the concept of aggregation of random variables is introduced, adapting the monotonicity and the boundary conditions by using stochastic orders. The choice of the stochastic order is crucial, with the usual stochastic order being the best alternative. A detailed study of the main families of aggregations of random variables and some classic properties of aggregation functions for the new concept is made.

Secondly, some particular cases of aggregations of random variables that are of special interest are considered. In particular, an induced linear fusion operator is considered in the context of mean estimation, some results about the asymptotic behavior of the optimal cumulative weights of L-estimators are proved, aggregations of random variables that order the inputs by means of their distribution functions are defined, properties of uniform random capacities are studied and a minimization problem that allows to define a particular type of approximation for aggregations of random variables is solved.

Posteriorly, a behavioral study of some aggregation functions, when their inputs are random variables, is made by using stochastic orders. First, it is concluded that some families of mean functions reduce the variability, with respect to the convex stochastic order, when applied to random vectors. It is also concluded that many variability measures take greater values when the associated random vector has more variability or less positive dependence. Although these results are intuitive, formal results in this regard cannot be found in the literature.

The notion of aggregation of random variables is extended to other random structures, with the cases of random elements on bounded posets, random vectors, stochastic orders and random sets being the ones studied more in detail. In addition to the construction of such notions, it is concluded that many properties of random vectors and random processes are preserved when the same usual aggregation function is applied componentwise or pointwise, respectively. For random sets, a location stochastic order is defined.

Finally, some of the developed methods are applied to real or simulated data. In particular, two prediction problems regarding temperature and humidity are addressed, the behavior of hypothesis tests related with the uniformity of random capacities, the dispersive stochastic order, the convex stochastic order and the supermodular order are studied and flexible-dimensional methods for estimating the mean of symmetric random variables are explored using simulation.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MATEMÁTICAS Y ESTADÍSTICA

Contents

Chapter 1

State of art and motivation

Contents

Classical aggregation functions are functions that take n real numbers over a real interval and return another value over the same interval, fulfilling monotonicity and some boundary conditions. These functions are usually used in data analysis, in problems such as the prediction or estimation of quantities of interest. From the perspective of Statistics, the available data can be considered to be observations of random variables. However, it is hard to find in the literature examples in this regard, in Aggregation Theory the data are usually considered to be just real numbers. This idea was the starting point for the study of aggregations of random variables, in which randomness is assumed to be present in both the input and the output of the aggregation function.

This very first chapter serves as an introduction to the thesis and explains in detail the motivation behind the idea of aggregation of random structures. Although

the idea has been considered tangentially in the literature [228, 269], to the best knowledge of the author, an exhaustive study of the notion has not been done until now. Thus, in the subsequent sections, some important initial questions are addressed. For instance,

- When should the probabilistic approach be considered?
- What are the benefits of such an approach?
- Why is this approach not the usual one in the literature?

In addition, the structure of the document is disclosed.

1.1 Probability in Aggregation Theory

As commented before, probability has been considered sporadically in the literature in Aggregation Theory, with different approaches. Firstly, it should be noted that the word *aggregation* is used in the literature with different meanings. For instance, in Probability Theory, there are some stochastic processes that consider aggregation with the meaning of the incorporation of elements to a cluster. This is the case of aggregation of Markov chains [291], aggregation of spaces areas in kinetic aggregation processes [10] or the random-time aggregation in partial adjustment models [178]. These cases, although with the term aggregation appearing, are not related to the approach followed in this work.

A reasonable amount of papers has been published regarding the so-called aggregation of probabilities. In this case, the inputs of the aggregations are probabilities of an event of interest or distribution functions. For instance, the weighted arithmetic mean of distribution functions is related to the mixture of random variables. Aggregation of probabilities has application in decision making [19, 91], geoscience [4] or energy consumption [316]. Some alternatives, such as aggregating density functions [216], mass functions [330, 331] or distribution functions [163] are considered. The reader is also referred to [67, 116, 251] for more information in this regard.

Some aggregations of probabilities can be linked to aggregations of random structures, see Section 3.1.3, but the concepts are quite different. In aggregation

of probabilities, even when working with distribution functions, there is not dependence structure between the possible random elements associated with the distributions. They have distribution functions as input for the aggregation, while, for instance, in the aggregation of random variables, the inputs are random variables. Of course, any random variable has an associated distribution function, but they are totally different mathematical notions. In addition, in many applied problems, the solution should be a particular random element, not a distribution. Moreover, many random structures, such as random graphs or random sets, do not have an associated distribution function.

Moving now to proper aggregation of random structures, one can find some examples in the literature, dealing mostly with random variables. Moreover, in general, it is considered only the case of usual aggregation functions with random inputs, that is, the composition of a measurable aggregation function A and a random vector $\vec{X}$, $A \circ \vec{X}$. It turns out that the result is a random variable and it will be a particular type of aggregation of random variables.

Most studies in this direction are devoted to determining properties of the aggregation function by means of the expectation of its value when the inputs are random variables. For instance, Chapter 10 in [149] is a clear example in this regard. It considers mainly the case of independent standard uniform random variables as the inputs of the aggregation function, and then computes the expectation of many indices and measures related to the aggregation. Some effort has been made to derive this type of properties for the Choquet and Sugeno integrals [117, 150, 193, 226, 227, 228, 230, 269].

Although previous papers are interesting, they do not provide a general theory of aggregation of random variables, they just consider some particular distribution for the inputs and compute expectations. There are many other cases of interest, considering non-uniform or dependent random variables. Moreover, the theory of aggregation of random variables developed in this thesis is wider than the one of the previous approach. As will be seen in Section 3.2, there are many aggregations of random variables that are not just the composition of usual aggregation functions and random vectors.

Finally, many examples of the use of aggregation functions applied to random variables can be found in Statistics. The arithmetic mean, the median and other

statistics are particular cases of aggregation functions. In general, almost any location measure [59] is an internal aggregation function (see Definition 2.8). However, again the study is often restricted to the composition of random vectors, having independent and identically distributed components in most of the cases, and location measures. Some additional parallelisms between Statistics and Aggregation Theory will be explored in Section 1.4.

Another important remark to make is that the general approaches given in [182, 197] do not include aggregations of random structures. They consider a partially ordered set $(S, \leq)$ with a minimum and maximum element and the componentwise order for vectors of elements in S. Even in the simplest case of random variables, having bounded spaces restricts the applicability, since many random variables have a non-bounded support. Moreover, when working with random variables, a stochastic order, see Section 2.3.5, should be considered. Stochastic orders are not antisymmetric, since they are defined in terms of the distribution of the random variables and there are different random variables with the same distribution. It should also be kept in mind that, when working with random vectors, it is usually not enough to have the ordering of the components for the ordering of the random vectors to be fulfilled.

1.2 When random structures must be considered?

The two main applied problems that justify the consideration of random structures in Aggregation Theory are prediction and estimation problems. These problems often have as initial data some measurements of a quantity of interest. If one follows the usual approach of Statistics [278], the measurements have been made over a random population. The values obtained are the result of several observations of such a quantity on a sample, which is a subset of the population. Then, the measurement process is, in fact, a random experiment and, therefore, the available data can be considered to have random behavior.

In prediction problems, several random quantities $X_1, \ldots, X_n, Y$ are considered. Then, the objective is to try to determine the value of Y when the values of $X_1, \ldots, X_n$ are given. If there is a positive dependence between $X_1, \ldots, X_n$ and Y, then it makes

sense to use an aggregation function A to fuse the values of $X_1, \ldots, X_n$, obtaining the prediction of Y as $A(X_1, \ldots, X_n)$. The positive dependence between $X_1, \ldots, X_n$ and Y is ensured, for instance, in many time series models [68, 243].

In this direction, aggregation functions are commonly used as ensembles. Suppose that one has n different prediction models for Y, each of them based on different techniques. In that scenario, it is reasonable to assume a positive dependence between the predictions of the prediction models. Therefore, predictions can be fused using an aggregation function to obtain a final prediction that overcomes the limitations of the initial prediction models [5, 284, 320]. The reader is referred to Section 4.1.1 for a detailed literature study of the use of aggregation functions as ensembles in time series.

In estimation problems, the most prominent example is the case of mean estimation. Suppose that there is a quantity of interest and one has a vector of observations of such a quantity perturbed by a random noise with mean 0. Then, estimating the mean of the associated random variable is equivalent to trying to determine the value of the quantity of interest. Many of the usual mean estimators are, in fact, aggregation functions [218, 278, 287, 288, 286]. The most remarkable examples are the arithmetic mean, the median, the maximum, the minimum or more general OWA operators.

It must be noticed that aggregation functions are used in other applications rather than prediction or estimation problems. For instance, t-norms are used to model intersection in Fuzzy Set Theory [100] and copulas are a powerful tool to model the dependence between random variables [118, 253], both being particular cases of aggregation functions. In these cases, a probabilistic approach is not adequate. Moreover, aggregation functions are often used in multi-criteria decision making [238], in which the opinion of experts sometimes cannot be seen as random values.

1.3 Why random structures must be considered?

As explained in the latter section, it is reasonable to consider randomness when dealing with prediction and estimation problems. However, it remains to clarify

the benefits of such an approach. Notice that by introducing random structures, the mathematical complexity increases, so its inclusion should be justified.

Focusing on random variables, they are measurable functions from a probability space $(\Omega, \mathscr{F}, P)$ to the real numbers endowed with the Borel σ-algebra, see Section 2.3.1. The probability space serves as an underlying structure that allows one to model the behavior of the population, not only the observed values. Moreover, the probability space is supposed to be the same for all the considered random elements, so it serves as a common structure that models their joint variation.

The first consequence of considering a probability approach is the notion of dependence. When just working with observations, it is not possible to take into account the possible relation between the inputs, whereas in many cases the values increase and decrease together. This is especially relevant in some applications such as the use of aggregation functions as ensembles in prediction problems, since, if the prediction models are reasonably good, they are going to have a strong and positive dependence.

Even when working with only one random variable, the variability cannot be properly modeled without random variables. Of course, one can compute some variability measures for a particular sample, but the results, without probability methods, cannot be extended to the population. Therefore, the information obtained might not be useful for a different observation of the same random variable.

In some applications, a training dataset is used to estimate some parameters of aggregation functions. Then, the fitted aggregation function is applied to new data to obtain a prediction or estimation of a quantity. Notice that, in this case, there are two sources of randomness. The first one is the randomness of the inputs of the aggregation function. The second one is the randomness associated with the estimation of the parameters, since they inherit the random behavior of the data that have been used to fit them. Without a probability approach, the possible relation between the fitted parameters and the input random variables, which appear, for instance, in time series, cannot be considered. This dependence can lead to some strange scenarios (see Example 3.21). Focusing on estimation, the probabilistic approach is crucial, since almost all properties of estimators cannot be defined without randomness (see Section 2.4.2).

Regardless of the latter considerations, there are some data analysis techniques

that are of interest and that, as being part of Statistics, require a probabilistic approach. For instance, the construction of confidence intervals, useful in order to give a set of possible values for a quantity, and hypothesis testing, which allows one to extend conclusions from the sample to the population, are only possible with random variables [278].

Finally, notice that another benefit of considering random variables is that there are properties of usual aggregation functions that only arise with a probabilistic approach, as can be seen in Chapter 10 in [149].

1.4 Parallelisms between Statistics and Aggregation Theory

As commented before, when aggregation functions are used in prediction and estimation problems, the considered approaches are usually equivalent to existing methods in Statistics. For instance, measures such as the Mean Squared Error or the Mean Absolute Error are used in both areas to measure the distance between the prediction/estimation and the real value. Since they share a common objective, Aggregation Theory and Statistics have independently developed some methods that are equivalent.

The most remarkable example is the use of the Ordered Weighted Averaging (OWA) operators, introduced in [327]. For computing these aggregation functions, the values of the inputs are ordered from the smallest to the greatest and then a convex linear combination is applied. From the side of Statistics, the approach of ordering samples and then computing a linear combination of the resulting values is a common procedure, see [104]. The resulting functions are called L-statistics. Some foundational papers can be found in the literature [218, 287, 288, 286], much older than the ones of OWA operators, which also give mathematical results on the behavior of such quantities. Of course, OWA operators can be seen as a particular case of L-statistics obtained when the linear combination is convex. For instance, the function given by

$$\frac{1}{2}\max(\vec{x}) + \frac{1}{2}\min(\vec{x}),$$

for any $\vec{x} \in \mathbb{R}^n$ is both an OWA operator and an L-statistic.

Another common point is the computation of quantities by minimizing a sort of dissimilarity measure between the input vector and the output value. For instance, it is known that the value y that minimizes

$$\sum_{i=1}^{n}(y-x_i)^2,$$

for any $\vec{x} \in \mathbb{R}^n$ is the arithmetic mean $y = \frac{1}{n}\sum_{i=1}^{n} x_i$. This approach is known in Aggregation Theory as aggregation based on penalties, see [75, 78, 304] and Section 6.6 in [149]. In Statistics, the same approach was considered before in the so-called extremum estimators [9, 11], which include, as particular cases, GMM estimators [166] and M-estimators [315].

Updating the result of the prediction or estimation is also a common point in some works on Statistics and Aggregation Theory. For instance, in Aggregation Theory, the notions of associativity [285] and recursivity [110] are really important to ease the computation of the aggregation. In Statistics, online estimation [175, 198, 332] also focuses on the update of a given estimation. The most remarkable example is the one concerning the arithmetic mean, since one has that

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_n + (n-1)\left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)}{n}.$$

It should be remarked that, even with all these points in common, there are many differences between Aggregation Theory and Statistics. As already commented before, Statistics considers a probability approach, while Aggregation Theory does not. On the other hand, the variety of functions that are considered in Aggregation Theory is wider than in Statistics, functions such as Choquet and Sugeno Integrals or t-norms are rarely used in Statistics.

Since some parts of Aggregation Theory and Statistics have common objectives and share methodological approaches, it is reasonable to try to combine both fields. For instance, some Statistical methods could be improved by using some involved aggregation functions or ideas from Aggregation Theory. In addition, Aggregation Theory, when used in data analysis, will also benefit from a probability

approach, as explained in the latter sections. Thinking about Statistics and Aggregation Theory as two different and independent branches of Mathematics can be a restriction when trying to develop better methods for data analysis.

1.5 Why random structures are not considered?

Although the benefits of considering random structures in data analysis are clear, as commented before, only a few papers and works consider this approach in Aggregation Theory. The reasons behind the absence of such a consideration are not simple, but some comments can be made.

Aggregation functions are used in many areas of Mathematics. However, an exhaustive study of such functions as a proper branch has been done mostly by the fuzzy set community. This fact can be verified by considering the authors of the most prominent books in the area [45, 49, 149], being all of them related to fuzzy sets. Fuzzy sets were introduced in [219] and generalize the concept of set by considering a membership function $\mu_A : S \to [0,1]$ that gives to each element $s \in S$ a degree of membership to A, $\mu_A(s)$. This degree of membership can be used to model imprecise data. For instance, if A is a set consisting of tall people, for a woman being 1.70 meters tall, it is not clear if the woman belongs to A or not. However, if a fuzzy set is considered, the woman can have a membership degree between 0 and 1 that reflects the subjectivity of the classification. Important types of functions in this regard are t-norms and t-conorms, which are used to define, respectively, the intersection and union of fuzzy sets [192]. It turns out that t-norms and t-conorms are increasing and fulfill the boundary conditions in 0 and 1. In [192], a generalization of them, aggregation functions, was introduced.

Therefore, it is quite natural for people from the fuzzy set community to be more interested in extending aggregation functions to deal with fuzzy concepts rather than to random structures. Many papers can be found with the first of the approaches [165, 275, 335]. Moreover, since the membership function takes values between $[0,1]$, just as a probability measure, a deep discussion about the interpretation about the necessity and semantics of fuzzy sets was made in the initial years. The theory of fuzzy sets was the subject of intense debate, see Section 2.4 in [52],

since some authors stated that Probability Theory was enough to achieve anything Fuzzy Set Theory could do [87, 212]. By some authors, see [57, 200], current consensus is that both theories deal with uncertainty. Fuzzy sets works with vagueness and Probability Theory with randomness, the first one being the imprecision of data, while the second the differences between the repetition of the same (random) experiment.

Although there is part of the community devoted to imprecise probability [237], fuzzy random variables [272] or to merge Probability Theory with fuzziness [281], this is not the case of researchers in Aggregation Theory, that might be reticent to use a fully probability approach in Aggregation Theory rather than a fuzzy one. Another possible reason is the expertise of researchers in Aggregation Theory. Most of them are computer scientists or mathematicians focused on artificial intelligence. Just a few experts on Probability Theory are nowadays working in the area, so it is possible that some involved notions of Probability Theory are not familiar to them.

It must be mentioned that, during the three years of preparation of this thesis, the approach of considering random structures in Aggregation Theory has been discussed in detail with many researchers in Aggregation Theory, with very positive feedback.

1.6 Structure and main objectives

The objective of this thesis is to state a theory of aggregation of random variables, study its utility and derive some applied methods to solve problems related to data analysis. In particular, three main objectives can be distinguished.

The definition and study of the concept of aggregation for different random structures. Adapting the monotonicity and the boundary conditions of usual aggregation functions to random structures is not straightforward. Therefore, it is necessary to construct a formal definition of such concepts for the considered random structures. The most usual ones, random variables, random vectors, stochastic processes and random sets require special attention. The proposed definitions should be studied in detail and the consequences of the chosen alternative for each case should be studied. Moreover, prominent families must be defined, focusing on

their semantics. For random structures such as random vectors and stochastic processes, the study of the preservation of some properties throughout the aggregation process should be addressed.

The definition of new aggregation operators, the study of their properties and their use in applied problems. Using the theoretical basis established in the latter point, some aggregation of random variables can be defined and used in applied problems. The definition of such aggregations should be stated and their mathematical properties should be studied. In addition, their applicability should be illustrated in some examples with real or simulated data.

The extension of the behavioral study of aggregation functions by means of random variables. Finally, using the probability approach, some properties of usual aggregations and related functions could be proved, following a similar idea as in Chapter 10 in [149]. In addition, the semantics and applicability of the proved results should be addressed.

The structure of the thesis is designed to answer the latter objectives. In Chapter 2, the basic notions that are needed for the development of the thesis are addressed, focusing mainly on Aggregation Theory, Probability Theory and Statistics. Then, Chapter 3 is devoted to the definition of aggregation of random variables, as well as to the study of some relevant families and properties of the introduced notion. Subsequently, several prominent examples of aggregations of random variables are given in Chapter 4. Its main properties are stated, with a focus on its applicability. Chapter 5 contains the results on stochastic inequalities considering aggregation functions and related functions, such as penalty functions and variability measures. The construction given in the first chapter is extended in Chapter 6, providing the definition of aggregations of random structures different from random variables. Chapter 7 is focused on the experimental part, including several applications of the defined aggregations of random variables and other developed techniques to real and simulated data.

Therefore, in order to clarify the structure of the thesis, in Figure 1.1 the relationship between different chapters and sections is detailed. For instance, it is recommended to read Section 4.1 before reading Section 7.6. Leaving aside the preliminaries, Chapter 3 serves as the basis for the rest of the chapters, so it is recommended to start from there.

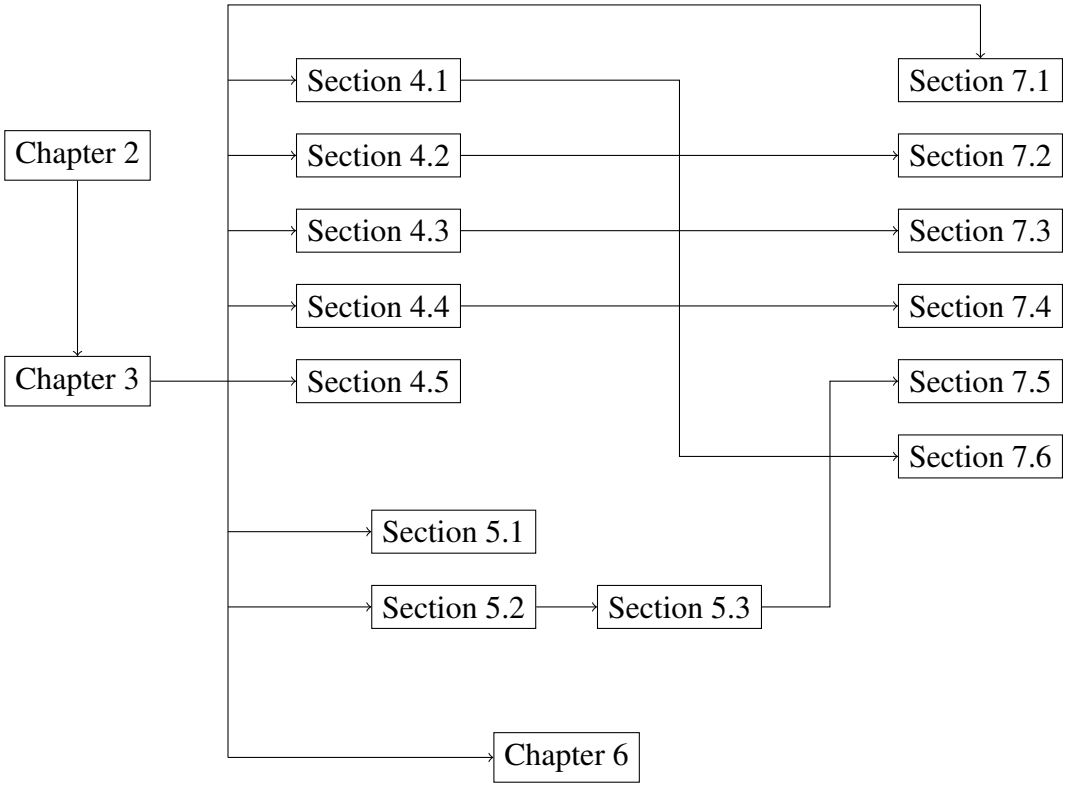

Figure 1.1: Scheme of the structure of chapters and sections of the thesis.

Regarding the three main objectives stated above, Chapters 3 and 6 are related to the definition of the concept of aggregation of random structures, Chapters 4 and 7 are associated with the definition of new aggregation operators and their use in applied problems. The results regarding the behavioral study of aggregation functions using random variables are provided in Chapter 5.

Chapter 2

Preliminaries

Contents

This section is devoted to explaining the basic concepts necessary for the correct development of the thesis. The considered topics are diverse, so many different definitions and notions have to be introduced. In particular, after some initial notions given in Section 2.1, the basics about Aggregation Theory, in Section 2.2, Probability Theory, in Section 2.3, and Statistics, in Section 2.4, are provided.

2.1 Some initial notions

Some initial notations and conventions have to be fixed before introducing more involved concepts. The symbols in brackets $\{\}$ are used to define sets. The cardinality of a set S will be denoted as $|S|$. The usual set operations such as the union $\cup$, the intersection $\cap$, the Cartesian product $\times$, the subtraction $\setminus$ of sets will be considered along the thesis with the usual definition. Sometimes, for the Cartesian product of n copies of the same set $S \times \cdots \times S$, the notation S^n will be considered. In addition,

the complementary of a subset of another set S, $B \subseteq S$, will be denoted as $\overline{B}$. Although most of the time the words *such that* will be explicitly written, sometimes the shortcut | will be used with the same meaning.

The set of real, integers and natural numbers will be denoted, respectively, as $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{N}$. If only positive real numbers, including 0, are considered, it will be denoted as $\mathbb{R}^+$. Given a set S, the set consisting of all the subsets of S, i.e. the parts of S, will be denoted by $\mathscr{P}(S)$. In general, any ordered finite collection of elements of a set S will be denoted as a vector $\vec{s} \in S$. In a vector space, the vectors $\vec{v}$ will be considered as column vectors, while their transposes $\vec{v}^t$ as row vectors.

The shortcut $[n]$ will be used to denote $\{1,\ldots,n\}$, with $n \in \mathbb{N}$. A permutation π over $[n]$ is a bijection from $[n]$ to $[n]$ and represents an order of the elements in $[n]$.

Many orders are going to be considered along the thesis. In many cases, the same symbol $\leq$ will be used with different meanings, whenever it does not lead to confusion. In general, it will refer to the usual order between real numbers or, when dealing with vectors, to the lattice or componentwise order of them.

Functions will be introduced as $f : S \to A$, where S and A are, respectively, the domain and the image of f. Given $B \subseteq S$, $f(B)$ will be used to denote the subset of A such that $a \in f(B)$ if there exists $b \in B$ with $f(b) = a$. The notation f^{-1} will be used for the pre-image or anti-image of f, that is, if $C \subseteq A$, then $f^{-1}(C)$ is a subset of S such that $s \in f^{-1}(C)$ if $f(s) \in C$. The symbol $\circ$ denotes the composition of functions.

The notation $(x_s, s \in S)$ is used to denote a collection of elements indexed using the elements of S. If $S = \mathbb{N}$, the collection is a sequence. The usual notion of sequence convergence is considered, denoting as $\lim_{n\to\infty} x_n$ the limit if it exists. If the elements of the sequence are real functions, $(f_s, s \in S)$, the sequence is said to be uniformly convergent over A if for any $\varepsilon \in \mathbb{R}^+$ there exists $m \in \mathbb{N}$ such that for any $k \in \mathbb{N}$ with $k > m$, it holds that $|f_k(t) - \lim_{n\to\infty} f_n(t)| < \varepsilon$. The limits of functions of real numbers, denoted as $\lim_{t\to t_0} f(t)$, are defined in the usual way.

Regarding the axioms for set theory, the here-considered approach is the usual one in modern mathematics, the Zermelo-Fraenkel axioms together with the axiom of choice (ZFC), see [289].

In the following sections, some brief notions about matrices, Topology and properties of functions are provided.

2.1.1 Matrices

A square matrix of dimension n $\left(M_{i,j}, i,j \in [n]\right)$, usually denoted just by M, is a collection of n^2 real values indexed over $[n]^2$. The trace $\mathrm{tr}(M)$ and the determinant $|M|$ are defined as usual. The inverse matrix of M, if it exists, is the (unique) matrix such that $M^{-1}M = MM^{-1} = I_n$, where I_n denotes the identity matrix of dimension n and the usual product of matrices is considered. The power of M^k is defined as the product of M by itself k times. In addition, the transpose of M, i.e. the interchange of rows and columns, will be denoted as M^t. The reader is referred to [319, 290] for more detailed information about matrices.

Notice that real vectors can be seen as matrices with just one column. Throughout the thesis, the notation $\vec{1}$ and $\vec{0}$ will be used to refer to the vector with all elements equal to 1 and 0, respectively. Given a squared matrix M, its infinite norm is defined as $||M||_\infty = \max_{i\in[n]} \sum_{j=1}^{n} |M_{i,j}|$.

Two particular types of matrices will be of interest. The first one is the class of positive semi-definite matrices. Recall that a squared matrix M is said to be symmetric if $M_{i,j} = M_{j,i}$ for any $i,j \in [n]$.

Definition 2.1 *[58] A squared symmetric matrix M of dimension n is said to be positive semi-definite if for any $\vec{x} \in \mathbb{R}^n$, $\vec{x}^t M \vec{x} \geq 0$.*

If the inequality is strict, then the matrix is positive definite. There are other characterizations of positive semi-definiteness by using eigenvalues or principal minors, see [58]. Another relevant type of matrix that will appear are the persymmetric matrices, which fulfill $M_{i,j} = M_{n+1-j,n+1-i}$ for any $i,j \in [n]$. The inverse of any invertible persymmetric matrix is also persymmetric [142]. The next result will be of interest for the proof of Proposition 4.11.

Lemma 2.2 *[155] Let A and B be two squared matrices such that A and $A-B$ are invertible. Then,*

$$(A-B)^{-1} = A^{-1} + A^{-1}B(A-B)^{-1},$$
$$(A-B)^{-1} = \sum_{k=0}^{\infty} \left(A^{-1}B\right)^k A^{-1}.$$

2.1.2 Topology

A topological space consists of a set S and a topology τ that includes some subsets of S, called open sets, which fulfill some particular properties. Formally,

Definition 2.3 *[248] Let S be a set. Then, $\tau \subseteq \mathscr{P}(S)$ is a topology of S if*

- $\emptyset, S \in \tau$,
- *For any arbitrary collection $(O_i, i \in I)$ with $O_i \in \tau$ for any $i \in I$, $\cup_{i\in I} O_i \in \tau$,*
- *For any $O_1, \ldots, O_n$ with $n \in \mathbb{N}$, $\cap_{i=1}^{n} O_i \in \tau$.*

The pair (S, τ) is called a topological space. Given a point $p \in S$, N is a neighborhood of p if there exists $O \in \tau$ such that $p \in O \subseteq N$. The complementary of any open set is called a closed set. The closure of a set B, $cl\,B$, is the smallest closed set that contains B. A particular type of sets, compact sets, will be of special importance.

Definition 2.4 *[248] Let (S, τ) be a topological space. Then, $K \subseteq S$ is said to be compact if, for any collection $(O_i, i \in I)$ such that $O_i \in \tau$ for every $i \in I$ and $K \subseteq \cup_{i\in I} O_i$, there exists a finite subset of I, $I_0 \subseteq I$, such that $K \subseteq \cup_{i\in I_0} O_i$.*

Therefore, a compact set is a set K such that, for any open cover of K, there always exists a finite subcover of K. A set $C \subseteq S$ is said to be connected if there do not exist two non-empty open sets O_1 and O_2 such that $O_1 \cap O_2 = \emptyset$ and $C = O_1 \cap O_2$ [248]. In $\mathbb{R}$, connected sets are bounded and unbounded intervals.

Some topologies can be constructed considering a distance (see Chapter 8 in [248]). If this is the case, the topological space is called metrizable. If the associated distance d implies that any Cauchy sequence, i.e. a sequence $(x_n, n \in \mathbb{N})$ for which for any $\varepsilon \in \mathbb{R}^+$ there exists $N \in \mathbb{N}$ such that for any $n, m \in \mathbb{N}$ with $n, m > N$ it holds $d(x_n, x_m) < \varepsilon$, is convergent, then the topological space is said to be completely metrizable. In addition, a topological space (S, τ) is said to be separable if there exists a countable set A such that $A \cap O \neq \emptyset$ for any $O \in \tau$. Separable and completely metrizable topological spaces are known as Polish spaces [141]. For

metrizable spaces, the closure of a set contains the limit of all convergent sequences contained in the set.

For the set $\mathbb{R}^n$, both the Euclidean distance, $d(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$, and the Manhattan distance, $d_M(\vec{x},\vec{y}) = \max_{i\in[n]} |x_i - y_i|$, generate the so-called usual topology of $\mathbb{R}^n$. This topological space is Polish [141].

2.1.3 Some properties of real functions

Consider a real function from the Cartesian product of the real line $f : \mathbb{R}^n \to \mathbb{R}$. Some properties of this type of functions are disclosed in this section. The here-presented properties can be found in many sources, for instance [295, 312]. Firstly, the terms positive and negative are going to be used in the non strict sense, i.e. 0 is both positive and negative. If 0 is excluded, then it will be used the terms strictly positive or strictly negative.

A real function is continuous if $\lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) = f(\vec{x}_0)$ for any $\vec{x} \in \mathbb{R}^n$. For functions $f : \mathbb{R} \to \mathbb{R}$, if the latter property holds for limits approaching from the left (right), the function is said to be left-continuous (right-continuous).

If it exists, the partial derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\frac{\partial}{\partial x_i} f(\vec{x}) = \lim_{h\to 0} \frac{f(\vec{x} + h\vec{e}_i) - f(\vec{x})}{h},$$

where $\vec{e}_i$ is the vector that takes the value 1 in the i-th position and 0 in the rest. If partial derivatives can be computed for any $\vec{x} \in \mathbb{R}^n$, it is said that the function is differentiable. If $f : \mathbb{R} \to \mathbb{R}$, the unique partial derivative will be denoted by f'. Similarly, the second partial derivatives $\frac{\partial^2}{\partial x_i \partial x_j} f(\vec{x})$ and $\frac{\partial^2}{\partial^2 x_i^2} f(\vec{x})$ are defined by considering two subsequent partial derivatives.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be increasing if $f(\vec{x}) \leq f(\vec{y})$ whenever $\vec{x} \leq \vec{y}$. The term strictly increasing will be reserved for the case $f(\vec{x}) < f(\vec{y})$. If f is differentiable, then it is increasing if and only if its partial derivatives are positive. Similarly, f is said to be (strictly) decreasing if $-f$ is (strictly) increasing.

In addition, a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $f(\lambda\vec{x} + (1-\lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$ for any $\vec{x},\vec{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$. If the latter inequality is strict, then f is said to be strictly convex. If, when a value for $n-1$ of the components of the vector is fixed, the resulting univariate function is convex, f is said to be componentwise

convex. If the second partial derivatives of f exists, then f is convex if and only if the Hessian matrix $H(\vec{x})$, defined as $H(\vec{x})_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} f(\vec{x})$ for any $i, j \in [n]$ is positive semi-definite for any $\vec{x} \in \mathbb{R}^n$. If $H(\vec{x})_{i,i}$ is positive for any $\vec{x} \in \mathbb{R}^n$ and $i \in [n]$, then f is componentwise convex. Similarly, f is said to be (strictly, componentwise) concave if $-f$ is (strictly, componentwise) negative.

A (permutation) symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as a function that satisfies $f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for any permutation $\pi : [n] \to [n]$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is supermodular if $f(\vec{x}) + f(\vec{y}) \leq f(\vec{x} \vee \vec{y}) + f(\vec{x} \wedge \vec{y})$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$, where $\vee$ and $\wedge$ denote, respectively, the componentwise maximum and minimum. If $-f$ is supermodular, then f is said to be submodular.

A contraction is a function $f : \mathbb{R}^n \to \mathbb{R}^n$ such that, for any $\vec{x}, \vec{y} \in \mathbb{R}^n$, it holds that $\sum_{i=1}^n (x_i - y_i)^2 \geq \sum_{i=1}^n (f(x_i) - f(y_i))^2$.

Finally, it is worth introducing the notation for the Riemann-Stieltjes integral. Given two real functions $f, g : I \to \mathbb{R}$ in a real interval I, such an integral is denoted as

$$\int_I f(x) dg(x).$$

This integral is defined in terms of the limit of sequences involving the sums

$$\sum_{i=1}^{n} f(x_i) \left(g(x_i) - g(x_{i-1}) \right),$$

where $x_0 \leq \cdots \leq x_n$ are the points associated with a partition of I. Notice that, if $g(x) = x$, then the result is the usual Riemann integral. The reader is referred to [74] for more information in this regard.

2.2 Aggregation Theory

Aggregation Theory is devoted to studying the fusion of information by means of particular functions called aggregation functions. The first ideas of aggregation were born related to the concept of averages or means. It is possible to find work about means by Cauchy in the 19th century [84], considering a mean to be a function that is bounded between the minimum and the maximum. More recently, Kolmogorov defined a mean as a continuous, increasing, symmetric, associative and

idempotent function [108, 196], which turns out to be the class of quasi-arithmetic means [73].

From another different perspective, aggregation functions also appear as a generalization of some operators that were relevant in Fuzzy Set Theory [234]. In this direction, the formal definition of aggregation function was proposed in [192], requiring the function to be increasing and to fulfill some boundary conditions. This general definition, which is the usual consideration in the present, includes functions that are used in very different topics such that Fuzzy Set Theory [190], decision making [229], data analysis [170] and dependence modeling [118, 253].

2.2.1 Aggregation functions

An aggregation function is typically referred to as a function that summarizes several values by a single number. In addition, any aggregation function should increase if all its arguments increase (monotonicity), and returns the supremum or the infimum when, respectively, the inputs are all the supremum or the infimum of the interval.

Throughout the thesis, I will be used to denote a closed (bounded or not) interval of the real line $\mathbb{R}$. The classical definition of aggregation function involves a bounded interval.

Definition 2.5 *[45] Let $I = [a,b]$ be a non-empty bounded real interval. A function $A : I^n \to I$ is said to be an aggregation function if*

1. *Is increasing,*
2. $A(a,\ldots,a) = a$,
3. $A(b,\ldots,b) = b$.

Conditions 2 and 3 are known, respectively, as the lower and upper boundary conditions. Many aggregation functions, such as t-norms, t-conorms, copulas, nullnorms or uninorms consider I to be the unit interval $[0,1]$. In fact, some approaches consider I to always be the unit interval, see [49].

However, sometimes it is necessary to aggregate values over unbounded intervals, typically the real line $\mathbb{R}$. This case is especially relevant in the approach of this thesis, since many important families of random variables have unbounded support. For instance, Gaussian and exponential random variables take values, respectively, in $\mathbb{R}$ and $\mathbb{R}^+$.

Definition 2.6 *[149] Let I be a non-empty real interval. A function $A : I^n \to I$ is said to be an aggregation function if*

1. *Is increasing,*
2. $\inf_{\vec{x}\in I^n} A(\vec{x}) = \inf I$,
3. $\sup_{\vec{x}\in I^n} A(\vec{x}) = \sup I$.

The boundary conditions for the unbounded case are not straightforward and deserve a detailed explanation. In the approach given in [149], it is considered the extended real line $\bar{\mathbb{R}} = [-\infty, \infty] = \mathbb{R} \cup \{-\infty, \infty\}$, where $-\infty$ and ∞ are two new elements that are added to $\mathbb{R}$. The usual order for real numbers is also extended to these new elements by stating $-\infty < x < \infty$ for any $x \in \mathbb{R}$.

With the latter consideration, the boundary conditions are well-defined even for unbounded intervals. They can be interpreted as the image of the aggregation having the same minimum (maximum) as the interval, if it exists, while not having a lower (upper) bound if the interval does not have a lower (upper) bound. Moreover, it is possible to rewrite such boundary conditions as follows.

2. For any $x \in I$, there exists $\vec{x} \in I^n$ such that $A(\vec{x}) \leq x$,
3. For any $x \in I$, there exists $\vec{x} \in I^n$ such that $A(\vec{x}) \geq x$,

which are trivially equivalent to the other ones using that A is increasing.

Finally, there are cases where the number of inputs is not fixed. For instance, the maximum can be defined for any finite number of values. In that direction, extended aggregations are considered.

Definition 2.7 *[149] Let I be a non-empty real interval. A function $A : \cup_{n\in\mathbb{N}} I^n \to I$ is said to be an extended aggregation function if any function $A^{(n)} : I^n \to I$ defined as $A^{(n)}(\vec{x}) = A(\vec{x})$ for all $\vec{x} \in I^n$ is an aggregation function.*

Notice that any extended aggregation can be seen as a collection of usual aggregation functions, one for each dimension of the inputs. However, typically the aggregation functions associated with an extended aggregation function are defined by the same procedure and share a common structure. Sometimes, for instance in the case of nullnorms and uninorms (see Definitions 2.18 and 2.19), they can be defined by using a recursive formula.

The notation A will be used for both aggregations and extended aggregations whenever it does not lead to confusion. Given an extended aggregation A, the aggregation functions associated with a fixed cardinality $n \in \mathbb{N}$ will be denoted as $A^{(n)}$ (just as in Definition 2.7). In addition, the convention of unary aggregation functions [149] will be considered, which states that $A^{(1)}(x) = x$ for any $x \in I$.

2.2.2 Properties of aggregation functions

In the literature, there are a huge number of properties that have been defined and studied for aggregation functions. Some of the typical properties have already been defined in Section 2.1.3, such as (permutation) symmetry or convexity, but others are particular to the theory of aggregation.

By the classification given in the main reference books [45, 49, 149], any aggregation function can be disjunctive, conjunctive or an average if, respectively, takes values smaller than the minimum, values greater than the maximum or takes values between the minimum and the maximum. If the aggregation is not included in any of the latter families, it is called mixed. For instance, copulas (see Section 2.3.2), are disjunctive and weighted arithmetic means are averages. The most important family in the approach given in this thesis is the case of averaging functions, also known as means. The property of being between the maximum and the minimum is usually called internality.

Definition 2.8 *[149] An aggregation function $A : I^n \to I$ is said to be internal if* $\min(\vec{x}) \leq A(\vec{x}) \leq \max(\vec{x})$ *for any* $\vec{x} \in I^n$.

It is well-known that, for aggregation functions, internality is equivalent to another relevant property, the idempotence. Intuitively, an aggregation function is

idempotent if the image of a constant vector is equal to the value of any of its components.

Definition 2.9 *[149] An aggregation function $A : I^n \to I$ is said to be idempotent if $A(x,\ldots,x) = x$ for any $x \in I$.*

Proposition 2.10 *[149] An aggregation function is internal if and only if it is idempotent.*

Idempotence, equivalently internality, is often required when using aggregation functions in applied problems related to fusion of prediction models, since it is reasonable to think that, if all models agree in one value, the aggregated prediction should be that value. Another property that is often relevant is to have the aggregations to have a good behavior when a linear transformation is made to the inputs. In general, one should expect coherent aggregation when there is a change in the units or a constant value is added to the fused information.

Definition 2.11 *[149] An aggregation function $A : I^n \to I$ is said to be interval scale invariant if $A(\mu\vec{1} + \lambda\vec{x}) = \mu + \lambda A(\vec{x})$ for any $\vec{x} \in I^n$, $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R}^+$ such that $\mu\vec{1} + \lambda\vec{x} \in I^n$.*

For example, the arithmetic mean and the maximum are interval scale invariant, whereas copulas are not. Notice that any of the latter properties can also be defined for any extended aggregation function by requiring the aggregation functions associated with the different cardinalities of the inputs to hold that property. In addition, there are some properties that are only defined for them, such as associativity.

Definition 2.12 *[149] An extended aggregation function $A : \cup_{n\in\mathbb{N}} I^n \to I$ is said to be associative if*

$$A^{(2)}\left(A^{(n)}(\vec{x}), A^{(m)}(\vec{y})\right) = A^{(n+m)}(\vec{x},\vec{y}),$$

for any $\vec{x} \in I^n$, $\vec{y} \in I^m$ and $n,m \in \mathbb{N}$.

Notice that any associative extended aggregation function can be computed by iteratively applying the associated aggregation function of dimension $n = 2$. For

instance, the maximum is associative since $\max(x_1,x_2,x_3) = \max(x_1,\max(x_2,x_3))$. Another related notion is the recursivity [110, 143] in which the applied function may change with the value of n. For instance, the arithmetic mean is recursive, since

$$A^{(n)}(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n-1} x_i + \frac{1}{n}x_n = \frac{n-1}{n}A^{(n-1)}(x_1,\ldots,x_{n-1}) + \frac{1}{n}x_n.$$

Another notion defined for extended aggregation functions is the existence or not of a neutral element. Roughly speaking, a neutral element is a value that can be ignored when computing the aggregated value.

Definition 2.13 *[149] Let $A : \cup_{n\in\mathbb{N}} I^n \to I$ be an extended aggregation function. Then, if there exists $e \in I$ such that $A(x_1,\ldots,x_{i-1},e,x_{i+1},\ldots,x_n) = A(x_1,\ldots,x_{i-1}, x_{i+1},\ldots,x_n)$ for any $n,i \in \mathbb{N}$ such that $i \leq n$ and $x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n \in I$, then e is said to be a neutral element of A.*

For instance, if the maximum of positive numbers is considered, then 0 is a neutral element, since $\max(0,x) = x$ for any $x \in \mathbb{R}^+$. The neutral element, if it exists, is unique [149] when $A^{(1)}(x) = x$ for any $x \in I$. On the other hand, there exists a type of elements, the annihilators, that have the opposite behavior. If the value is included among the inputs, then the output value is equal to the annihilator element.

Definition 2.14 *[149] Let $A : \cup_{n\in\mathbb{N}} I^n \to I$ be an extended aggregation function. Then, if $A(\vec{x}) = a$ for any $\vec{x} \in \cup_{n\in\mathbb{N}} I^n$ such that $a \in \{x_1,\ldots,x_n\}$, then a is said to be an annihilator element of A.*

An example of an annihilator element is 0 in the geometric mean. Similarly as with neutral elements, if there exists an annihilator element, it is unique [149]. However, it is possible to have both a neutral and an annihilator element for the same aggregation. For instance, the maximum over the unit interval has as neutral element 0 and as annihilation element 1.

2.2.3 Families of aggregation functions

In the development of this thesis, some particular families of aggregation functions will appear. This section is devoted to introducing them and providing their main properties. Many aggregation functions are defined using a weighting vector, which is just a vector with positive components that sum 1. More precisely, a vector $\vec{w} \in \mathbb{R}^n$ is said to be a weighting vector if $\vec{w} \in [0,1]^n$ and fulfills $\sum_{i=1}^{n} w_i = 1$. In some parts of this thesis, such as Sections 4.1 and 4.2, positivity of the weights will be dropped. In that case, the term generalized weighting vector will be used. One of the families, which definition implies a weighting vector, are the weighted quasi-arithmetic means. They are obtained by generalizing the arithmetic mean, considering a function h that changes the scale of the inputs and a weighting vector.

Definition 2.15 *[73] An aggregation function $A : I^n \to I$ is said to be a weighted quasi-arithmetic mean if it can expressed as,*

$$A(\vec{x}) = h^{-1}\left(\sum_{i=1}^{n} w_i h(x_i)\right), \forall \vec{x} \in I^n,$$

with $h : I \to \mathbb{R}$ being a real-valued strictly monotone function and $\vec{w}$ being a weighting vector.

Weighted quasi-arithmetic means are also known as the Kolmogorov-Nagumo means [73]. They include, among others, weighted arithmetic means, weighted geometric means, weighted harmonic means and weighted power means, which have, respectively, the following expressions.

$$WAM(\vec{x};\vec{w}) = \sum_{i=1}^{n} w_i x_i, \quad WGM(\vec{x}) = \prod_{i=1}^{n} x_i^{w_i},$$

$$WHM(\vec{x};\vec{w}) = \frac{1}{\sum_{i=1}^{n} \frac{w_i}{x_i}}, \quad WPM(\vec{x};\vec{w},p) = \left(\sum_{i=1}^{n} w_i x_i^p\right)^{\frac{1}{p}},$$

with $p \in [\infty, -\infty]$ and the inputs $\vec{x}$ restricted to $\mathbb{R}^+$ in all cases but the weighted arithmetic mean. When p is $-\infty$ or ∞, the aggregation is, respectively, the minimum and the maximum.

Another family that uses a weighting vector but also considers the ordering of the inputs are the Ordered Weighted Averaging (OWA) operators.

Definition 2.16 *[327] An aggregation function $A : I^n \to I$ is said to be an Ordered Weighted Averaging if it can be expressed as,*

$$A(\vec{x}) = \sum_{i=1}^{n} w_i x_{\pi(i)},$$

where $\pi : [n] \to [n]$ denotes a permutation such that $x_{\pi(1)} \leq \cdots \leq x_{\pi(1)}$ and $\vec{w}$ is a weighting vector.

Any OWA operator will be denoted by $OWA(\vec{x};\vec{w})$. When $\vec{w} = (1,0,\ldots,0)$, the OWA operator behaves as the minimum and, when $\vec{w} = (0,\ldots,0,1)$, as the maximum. As has been discussed in Section 1.4, it has a deep relation with convex linear combinations of order statistics, which will be introduced in Section 2.3.2.5.

With an adequate choice of the weights, they are robust to outliers. In particular, trimmed means, which are defined as the average of the central values of the ordered sample, are particular types of OWA operators. Moreover, the median, both for an even and an odd number of inputs, is an OWA operator.

In other scenarios, the arguments of the aggregation are ordered by means of an auxiliary vector. This is the case of the Induced Ordered Weighted Averaging.

Definition 2.17 *[328] Let $\vec{w} \in \mathbb{R}^n$ be a weighting vector. Consider the permutation $\pi_{\vec{y}} : [n] \to [n]$ such that $\pi_{\vec{y}}(\vec{y})_1 \leq \cdots \leq \pi_{\vec{y}}(\vec{y})_n$ and, if there is any draw in $\vec{y}$, replace the associated values of $\vec{x}$ by their average. Then, the Induced Ordered Weighted Averaging (IOWA) has the following expression:*

$$IOWA(\vec{x},\vec{y};\vec{w}) = \sum_{i=1}^{n} w_i \pi_{\vec{y}}(\vec{x})_i.$$

Notice that the IOWA operator is an aggregation function only when the vector $\vec{y}$ is fixed, since a change of the order of the variables could make the monotonicity not hold. However, a correct choice of the vector that induces the order can be interesting from an applied point of view. The reader is referred to Section 4.1.1 for more information about the applicability of this type of functions in time series forecasting.

Both weighted quasi-arithmetic means and OWA operators contain, as special cases, the maximum and the minimum. Other families, nullnorms and uninorms,

also generalize the maximum and the minimum. Rather than an explicit expression, their definitions consider the interval I to be the unit interval $[0,1]$, a flexible number of inputs and impose some properties.

Definition 2.18 *[149] An extended aggregation function $A : \cup_{n\in\mathbb{N}}[0,1]^n \to [0,1]$ is said to be a nullnorm if it is symmetric, associative and has an annihilator element $a \in [0,1]$.*

Definition 2.19 *[149] An extended aggregation function $A : \cup_{n\in\mathbb{N}}[0,1]^n \to [0,1]$ is said to be an uninorm if it is symmetric, associative and has a neutral element $e \in [0,1]$.*

Nullnorms and uninorms will be denoted, respectively, by N and U. If the annihilator element or the neutral element must be explicitly stated, the notation N_a and U_e will be used. Notice that, as already introduced in Section 2.2.2, annihilator and neutral elements are unique.

The structure of nullnorms and uninorms has a very strong relation with t-norms and t-conorms. Since the general case will not be of interest in this thesis, the reader is referred to [149] for more information in this regard. The particular case that will be considered is the one of idempotent nullnorms and uninorms, which have easier characterizations.

Proposition 2.20 *[100] Let $N : \cup_{n\in\mathbb{N}}[0,1]^n \to [0,1]$ be an idempotent nullnorm with annihilator element $a \in [0,1]$. Then,*

$$N(x,y) = \begin{cases} \max(x,y) & \text{if } x,y < a, \\ \min(x,y) & \text{if } x,y > a, \\ a & \text{elsewhere.} \end{cases}$$

Proposition 2.21 *[100] Let $U : \cup_{n\in\mathbb{N}}[0,1]^n \to [0,1]$ be an idempotent uninorm with neutral element $e \in [0,1]$. Then, there exists a decreasing function $g : [0,1] \to [0,1]$ with $g(e) = e$ such that:*

$$U(x,y) = \begin{cases} \min(x,y) & \text{if } y < g(x), \\ \min(x,y) \text{ or } \max(x,y) & \text{if } y = g(x), \\ \max(x,y) & \text{if } y > g(x). \end{cases}$$

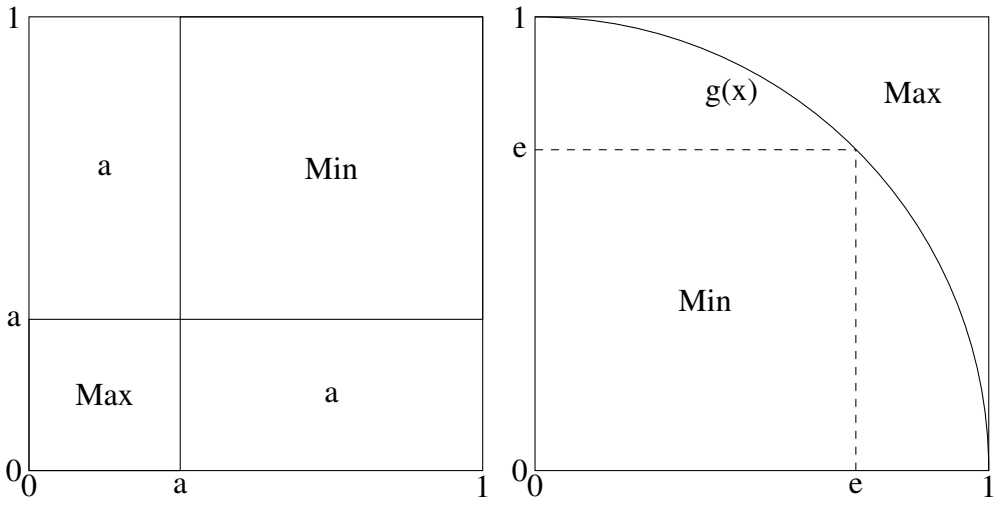

Figure 2.1: Structure of an idempotent nullnorm with annihilator element a (left) and an idempotent uninorm with neutral element e (right).

It should be noted that, while the expression in Proposition 2.20 is a characterization, the expression given by Proposition 2.21 it is not. That is, not every decreasing function $g : [0,1] \to [0,1]$ with $g(e) = e$ leads to an idempotent uninorm.

Although the latter results only give the structure when the number of inputs is 2, the associativity allows to construct the value of the nullnorm or uninorm just by the recursive application of the bidimensional case. For instance, for dimension 3 one can write $N(x_1,x_2,x_3) = N(N(x_1,x_2),x_3)$ and $U(x_1,x_2,x_3) = U(U(x_1,x_2),x_3)$. An example of such structures can be found in Figure 2.1.

2.2.4 Capacities and related integrals

This section is devoted to introducing other relevant families of aggregation functions, the Choquet and Sugeno Integrals, and the main concepts that are necessary to define them, capacities. Moreover, many properties and measures related to capacities will be provided, since they will be relevant in Section 4.3.

A capacity is a generalization of the concept of measure in which the additivity is replaced by monotonicity with respect to the inclusion order. Throughout the thesis, only finite capacities associated with the set $[n] = \{1,\ldots,n\}$ and the σ-algebra

(see Definition 2.34) consisting of all the subsets of $[n]$, $\mathscr{P}([n])$, will be considered.

Definition 2.22 *[48, 146] A function $\mu : \mathscr{P}([n]) \to [0,1]$ is said to be a capacity (of dimension n) if the following conditions are fulfilled:*

- $\mu(\emptyset) = 0$,
- $\mu(A) \leq \mu(B)$ *for any* $A, B \subseteq [n]$ *such that* $A \subseteq B$.

Throughout the thesis, all capacities will be normalized, i.e. $\mu([n]) = 1$. Capacities are also known in the literature as fuzzy measures [48] or as particular types of games [146]. The relaxation of additivity allows one to consider scenarios in which the elements that are being considered can interact. In some cases, it is convenient to represent capacities in the so-called Mobius representation.

Definition 2.23 *[48, 146] Let $\mu : \mathscr{P}([n]) \to [0,1]$ be a capacity. Then, the Mobius representation of μ is a function $M_\mu : \mathscr{P}([n]) \to [0,1]$ such that*

$$M_\mu(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \mu(B),$$

for any $A \subseteq [n]$.

This representation characterizes the capacity, since two Mobius representations are equal if and only if the capacities are the same [48]. From the Mobius representation, one can obtain the capacity by computing

$$\mu(A) = \sum_{B \subseteq A} M_\mu(B),$$

for any $A \subseteq [n]$.

In the following, two descriptive quantities that will be of interest in the subsequent study are introduced. Firstly, the Orness of the Choquet integral (see Definition 2.27), measures how far the aggregation is from the minimum depending on the capacity.

Definition 2.24 *[48] Let $\mu : \mathscr{P}([n]) \to [0,1]$ be a capacity. Then, the Orness of C_μ, $O(C_\mu)$, is defined as*

$$O(C_\mu) = \frac{1}{n-1} \sum_{A \subset [n]} \frac{(n-|A|)!|A|!}{n!} \mu(A).$$

Secondly, the nonmodularity index measures the deviation from the modular property (see [48]) for each subset.

Definition 2.25 *[48] Let $\mu : \mathscr{P}([n]) \to [0,1]$ be a capacity. Then the nonmodularity of $A \subseteq [n]$, $d_\mu(A)$, is defined as*

$$d_\mu(A) = \mu(A) - \frac{1}{|A|} \sum_{i \in A} \left(\mu(\{i\}) + \mu(A \backslash \{i\})\right).$$

Another useful concept is the notion of linear extension of a capacity. Given a capacity $\mu : \mathscr{P}([n]) \to [0,1]$, an ordering of the elements of $\mathscr{P}([n])$, $A_1, \ldots, A_{2^n}$ such that $\mu(A_1) \leq \cdots \leq \mu(A_{2^n})$ is said to be a linear extension. From the definition of capacities, there exist restrictions about the possible linear extensions. For instance, for $n = 2$, if there are no repeated values, only the cases $\mu(\emptyset) \leq \mu(\{1\}) \leq \mu(\{2\}) \leq \mu(\{1,2\})$ and $\mu(\emptyset) \leq \mu(\{2\}) \leq \mu(\{1\}) \leq \mu(\{1,2\})$ are possible. The number of linear extensions increases at a very high rate with n [96, 146]. There are many families of capacities, most of them related to conditions that ease their structure, thus restricting the study to a particular family of capacities simplifies computations. They will be of interest in Section 4.3.3.

Definition 2.26 *[48] Let $\mu : \mathscr{P}([n]) \to [0,1]$ be a capacity. Then,*

- *If $\mu(A) = \mu(B)$ for each $A, B \subseteq [n]$ such that $|A| = |B|$, μ is said to be symmetric,*
- *If $\mu(A) \leq \mu(B)$ for each $A, B \subseteq [n]$ such that $|A| \leq |B|$, μ is said to be balanced,*
- *If $\mu(A) = 1 - \mu(\bar{A})$ for any $A \subseteq [n]$, μ is said to be auto-dual,*
- *If $\mu\left(\cup_{i=1}^m A_i\right) \geq \sum_{B \subseteq [m]} (-1)^{|B|+1} \mu\left(\cap_{i \in B} A_i\right)$, for each $A_1, \ldots, A_m \subseteq [n]$ with $1 < m \in \mathbb{N}$, μ is said to be a belief measure,*
- *If $\mu(A \cup B) = \max(\mu(A), \mu(B))$ for any $A, B \subseteq [n]$, μ is said to be a possibility measure,*
- *If $\mu(A) \in \{0,1\}$ for any $A \subseteq [n]$, μ is said to be a 0-1 capacity.*

In Aggregation Theory, the main application of capacities is their use in the so-called Choquet and Sugeno integrals. These aggregation functions take into account the possible interaction between the different inputs and are widely used in many applications [147, 151, 183].

Definition 2.27 *[149] Let μ be a capacity of dimension n. Then, the Choquet integral of $\vec{x}$ with respect to μ is defined as*

$$C_\mu(\vec{x}) = \sum_{i=1}^{n} \left(x_{\pi(i)} - x_{\pi(i-1)}\right) \mu(A_{\pi(i)}),$$

with π being a permutation of $[n]$ such that $x_{\pi(1)} \leq \cdots \leq x_{\pi(n)}$, with the convention $x_{\pi(0)} = 0$ and $A_{\pi(i)} = \{\pi(i), \ldots, \pi(n)\}$.

Definition 2.28 *[149] Let μ be a capacity of dimension n. Then, the Sugeno integral of $\vec{x}$ with respect to μ is defined as*

$$S_\mu(\vec{x}) = \max_{i \in [n]} \left(\min\left(x_{\pi(i)}, \mu(A_{\pi(i)})\right)\right),$$

with π being a permutation of $[n]$ such that $x_{\pi(1)} \leq \cdots \leq x_{\pi(n)}$ and $A_{\pi(i)} = \{\pi(i), \ldots, \pi(n)\}$.

2.2.5 Another notions of aggregation

Aggregation functions, while initially defined for real values, have been extended and adapted to many different scenarios. For instance, the aggregation of different mathematical objects such as graphs [122], elements in bounded lattices [182], vectors [235] or fuzzy sets [311] have been considered in the literature.

In other scenarios, the concept has been generalized by relaxing the condition of monotonicity [47, 322]. Moreover, there are also some papers dealing with different philosophies among the concept of aggregation [241, 266].

2.2.5.1 Weak monotonicity

In some cases, the monotonicity of aggregation functions is a strong restriction that reduces applicability. In this direction, some papers have shown that relaxing the

monotonicity for some families of aggregation functions can improve the results of classical aggregations for some tasks. For instance, the reader is referred to the generalization of the Choquet Integral, see [114], and its applicability to classification [220, 221] or image analysis [112], or to the use of t-norm based pre-aggregations in Knowledge-Based Systems [308]. The usual consideration is to replace monotonicity with directional monotonicity.

Definition 2.29 *[47] Let $A : I^n \to I$ be a function and $\vec{r} \in \mathbb{R}^n$. A is said to be directionally monotone with respect to $\vec{r}$ if*

$$A(\vec{x} + c\vec{r}) \geq A(\vec{x}),$$

for any $\vec{x} \in I^n$ and any $c \in \mathbb{R}^+$ such that $\vec{x} + c\vec{r} \in I^n$.

A particular case is weak monotonicity, which is directional monotonicity with respect to the vector of ones $\vec{1} \in \mathbb{R}^n$ [322]. Trivially, any interval scale invariant function is weakly increasing. The notion of pre-aggregation function follows the same idea as the definition of aggregation function, but replacing monotonicity by directional monotonicity.

Definition 2.30 *[76] A function $PA : [0,1]^n \to [0,1]$ is said to be a pre-aggregation function if*

1. *There exists $\vec{r} \in \mathbb{R}^n$ such that PA is directionally monotone with respect to $\vec{r}$,*
2. *$PA(0,\ldots,0) = 0$,*
3. *$PA(1,\ldots,1) = 1$.*

Of course, the latter definition can be easily adapted to any real interval I, as well as to inputs of different size, just by considering the same ideas as in Section 2.2.1. Any aggregation function is a pre-aggregation function for any positive vector $\vec{r} \in \mathbb{R}^{+n}$. The properties introduced in Section 2.2.2 can be extended for pre-aggregation functions.

For instance, the function $PA : \mathbb{R}^2 \to \mathbb{R}$ given by $PA(x_1, x_2) = -0.1x_1 + 1.1x_2$ is a pre-aggregation function that is increasing with respect to $\vec{r} = (1,1)$. In addition,

it is idempotent and interval scale invariant. However, it is not internal. Notice that Proposition 2.10 holds for aggregations, but not for pre-aggregation functions.

A particular case that will be of interest is the case of the OWA operator with possibly negative weights, for which the set of directions for which the function is directionally increasing can be found in Proposition 4.3 in [50].

Proposition 2.31 *[50] Let $PA : I^n \to I$ be a pre-aggregation function defined as*

$$PA(\vec{x}) = \max\left\{0, \min\left\{\sum_{i=1}^{n} w_i x_{\pi(i)}\right\}\right\},$$

where $\pi : [n] \to [n]$ denotes the permutation such that $x_{\pi(1)} \leq \cdots \leq x_{\pi(1)}$ and $\vec{w} \in \mathbb{R}^n$ a vector such that $\sum_{i=1}^{n} w_i = 1$. Then, PA is directionally increasing with respect to $\vec{r}$ if $r_i \geq 0$ for any $i \in [n]$ and

$$\sum_{i=1}^{n} r_{\sigma(i)} w_i \geq 0,$$

for any possible permutation $\sigma : [n] \to [n]$.

Notice that the original result also requires $\sum_{i=1}^{n} r_i = 1$, but it is clear that multiplying the values of $\vec{r}$ by a positive constant does not change the sign of the components or the sign of the associated sums.

2.2.5.2 Aggregation on bounded posets

Another technique that can be used to generalize aggregation functions is to change the type of elements that are being aggregated. Since monotonicity is defined in terms of the usual order between vectors, it is natural to consider a general ordered set.

Definition 2.32 *[102] Let S be a set and let $\leq$ be a binary relation. It is said that $\leq$ is a total order if the following conditions are fulfilled for any $a, b, c \in S$:*

1. *(Reflexivity) $a \leq a$,*
2. *(Antisymmetry) $a \leq b,\ a \leq b \implies a = b$,*
3. *(Transitivity) $a \leq b,\ b \leq c \implies a \leq c$,*

4. *(Completeness)* $a \not\leq b \implies b \leq a$.

The pair $(S, \leq)$ is said to be a totally ordered set. However, in many cases, the binary relation lacks completeness, all elements cannot be pairwise compared. A binary relation that fulfills conditions 1, 2 and 3 is called a partial order [102] and $(S, \leq)$ is said to be a partially ordered set or poset. A subset $U \subseteq S$ is said to be an upper (lower) set if for any $a, b \in S$ such that $a \in U$ and $b \geq (\leq) a$, then $b \in U$.

A poset $(S, \leq)$ is said to be bounded if there exist two elements $\mathbf{0}, \mathbf{1} \in S$ such that $\mathbf{0} \leq x \leq \mathbf{1}$ for any $x \in S$. They are denoted by $(S, \leq, \mathbf{0}, \mathbf{1})$. By slightly adapting the monotonicity and the boundary conditions, aggregation functions can be defined for bounded posets.

Definition 2.33 *[197] Let $(S, \leq, \mathbf{0}, \mathbf{1})$ be a bounded poset. A function $A : S^n \to S$ is said to be an aggregation function if*

1. $A(x_1, \ldots, x_n) \leq A(y_1, \ldots, y_n)$ *for any* $x_1, \ldots, x_n, y_1, \ldots, y_n \in S$ *such that* $x_i \leq y_i$ *for any* $i \in [n]$,
2. $A(\mathbf{0}, \ldots, \mathbf{0}) = \mathbf{0}$,
3. $A(\mathbf{1}, \ldots, \mathbf{1}) = \mathbf{1}$.

2.3 Probability Theory

Probability Theory is a branch of Mathematics that is devoted to study randomness, including random events, random elements and probability distributions. It is the basis of Statistics and it uses concepts from other areas such as measure theory, functional analysis and topology.

One of the first known works on Probability Theory is due to Al-Kindi in the ninth century, using frequency analysis applied to break ciphers [3]. More recently, the first book devoted to Probability Theory is *De Rationiis in Ludo Aleae* in 1657, written by Christian Huygens and mainly based on the works of Pascal and Fermat on problems related to tossing coins or throwing dice [109]. After some years, Jacob Bernoulli published *Ars Conjectandi*, proving fundamental results for Probability

Theory such that the law of big numbers [294]. One of the first attempts to extend Probability Theory beyond games of chance such as tossing coins or card games was made by Pierre Simon Laplace in 1812 with the book *Théorie Analytique des Probabilités* [205], including applications to life insurance and mortality. However, was Andréi Nikoláyevich Kolmogórov the one who established the modern theory of probability in his book entitled *Foundations of the Probability Theory* [195] in 1933. He gave a mathematical structure, the probability space, that allows one to have well-defined measurements of random experiments, known as random elements. This approach permits to work with variables instead of just working with probability distributions.

In the last century, Probability Theory has been extended in many directions with the appearance of a huge number of branches that focus on different aspects of probability. In this section, the concepts about Probability Theory that will be necessary for the presented results are provided. Although some of the notions are widely known, they will be introduced in order to have a self-contained document and to fix the notation. In particular, basic notions about probability spaces and random elements (Subsection 2.3.1), random vectors and copulas (Subsection 2.3.2), stochastic processes (Subsection 2.3.3), random sets (Subsection 2.3.4) and stochastic orders (Subsection 2.3.5) are provided.

2.3.1 Probability spaces and random elements

The axiomatic formulation of Probability Theory considers probability spaces as the main concept. Given a set Ω, often called the sample space, the set of subsets for which the probability can be computed is called the σ-algebra of the probability space. It is defined as follows.

Definition 2.34 *[23] Let Ω be a set and $\mathscr{F} \subseteq \mathscr{P}(\Omega)$ be a set of subsets of Ω. $\mathscr{F}$ is said to be a σ-algebra if the following conditions are fulfilled:*

- $\Omega \in \mathscr{F}$,
- *If $B \in \mathscr{F}$, then $\overline{B} \in \mathscr{F}$,*
- *If $B_1, B_2, \cdots \in \mathscr{F}$, then $\left(\cup_{i=1}^{\infty} B_i\right) \in \mathscr{F}$.*

The set of all subsets of Ω, denoted as $\mathbb{P}(\Omega)$, is an example of a simple σ-algebra. The pair $(\Omega, \mathscr{F})$ is called a measurable space. The elements of $\mathscr{F}$ are usually referred to as the measurable sets of the measurable space. Sometimes, it is necessary to construct a σ-algebra that contains some particular collection of sets. In this regard, the next result shows that this can be done univocally.

Proposition 2.35 *[23] Let $\mathscr{C}$ be a set of subsets of Ω. Then, there exists a unique σ-algebra over Ω, $\mathscr{F}$, such that $\mathscr{C} \subseteq \mathscr{F}$ and $\mathscr{F}$ is minimal with respect to this property.*

This σ-algebra, denoted by $\sigma(\mathscr{C})$, is known as the σ-algebra generated by $\mathscr{C}$. More precisely, $\sigma(\mathscr{C})$ can be constructed by considering countable unions, intersections and complements of elements in $\mathscr{C}$. One of the most prominent examples is the Borel σ-algebra, which is generated by the open sets of a given topology. In addition to the σ-algebra, any probability space has a probability measure.

Definition 2.36 *[278] Let $(\Omega, \mathscr{F})$ be a measurable space. A function $P : \mathscr{F} \to [0,1]$ is said to be a probability measure if it fulfills the following properties:*

- $P(\Omega) = 1$,
- *If $B_1, B_2, \ldots$ is a pairwise disjoint sequence of measurable sets, then*

$$P\left(\cup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P\left(B_i\right).$$

The trio $(\Omega, \mathscr{F}, P)$ is known as a probability space. There are many different probability spaces, some of them being very restrictive. For instance, if Ω is finite, it is not possible to define continuous random variables. In general, a desirable property is completeness.

Definition 2.37 *[23] A probability space $(\Omega, \mathscr{F}, P)$ is said to be complete if for any $B_0 \subseteq \Omega$ such that there exists $B \in \mathscr{F}$ that fulfills $B_0 \subseteq B$ and $P(B) = 0$, $B_0 \in \mathscr{F}$.*

Any probability space $(\Omega, \mathscr{F}, P)$ can be extended to a complete probability space $(\Omega, \mathscr{F}_0, P_0)$ such that $\mathscr{F} \subseteq \mathscr{F}_0$ and $P(B) = P_0(B)$ for any $B \in \mathscr{F}$ [23]. Otherwise stated, all probability spaces will be complete. One of the most used probability spaces is the unit interval with the Lebesgue measure, see [66]. In this case, $\Omega = [0,1]$ and the probability measure is the Lebesgue measure λ, defined as

$$\lambda(B) = \inf \left\{ \sum_{k=1}^{\infty} \text{vol}(C_k) \;\middle|\; C_k \text{ is an open interval } \forall k \in \mathbb{N} \text{ and } B \subseteq \cup_{k=1}^{\infty} C_k \right\},$$

where vol stands for the volume. The σ-algebra consists of sets $B \subseteq [0,1]$ such that $\lambda(A) = \lambda(A \cap B) + \lambda(A \cap \overline{B})$ for any $A \subseteq [0,1]$.

Given a probability space, random elements are functions from the probability space to another measurable space that fulfills a property related to the involved σ-algebras. Recall that, given a function $f : S_1 \to S_2$, f^{-1} denotes the pre-image, that is, $f^{-1}(B) = \{x \in S_1 \mid f(x) \in B\}$.

Definition 2.38 *[60] Let $(\Omega, \mathscr{F}, P)$ be a probability space and let $(S, \mathscr{G})$ be a measurable space. Then, a function $X : \Omega \to S$ such that $X^{-1}(B) \in \mathscr{F}$ for any $B \in \mathscr{G}$ is called a random element.*

The latter property is known in measure theory as measurability. When the random element is constant, it is said to have a degenerate distribution and is generally denoted by its value in the set S.

The most prominent example of random elements, random variables, is obtained by considering $(S, \mathscr{G})$ as the real numbers with the Borel σ-algebra associated with the usual topology of real numbers. In this direction, it is known that, given two random variables X and Y, its sum and its product are measurable [93]. Moreover, the set $\{\omega \in \Omega \mid X(\omega) \leq Y(\omega)\}$ is also measurable [93].

Notice that any random element X induces a probability measure P_X in the measurable space $(S, \mathscr{G})$. In particular, $P_X(B) = P(X^{-1}(B))$ for any $B \in \mathscr{G}$. This probability measure is known as the distribution of X. In the following, $P_X(B)$ will simply be denoted as $P(X \in B)$. If $P_X(\{s\}) > 0$, the element $s \in S$ is said to be a probability mass point of X.

Two random elements are said to have the same distribution, often denoted by $X =_{st} Y$ (in an abuse of notation), if $P(X \in B) = P(Y \in B)$ for any $B \in \mathscr{G}$. A

stronger property is to be almost surely equal, denoted by $X =_{a.s.} Y$, which means that $P(X = Y) = P(\{w \in \Omega \mid X(w) = Y(w)\}) = 1$. Of course, $X =_{a.s.} Y \implies X =_{st} Y$.

In general, to check if two random elements have the same distribution, it is necessary to check the probability for all measurable subsets. However, in some cases, a shortcut can be used.

Definition 2.39 *[119] Let S be set and $\mathscr{S} \subseteq \mathscr{P}(S)$. Then, $\mathscr{S}$ is said to be a π-system if*

- $\Omega, \emptyset \in \mathscr{S}$,
- $A \cap B \in \mathscr{S}$ *for any* $A, B \in \mathscr{S}$.

The following is a slight modification of Theorem A.1.5. in [119].

Theorem 2.40 *Let $(\Omega, \mathscr{F}, P)$ be a probability space, $\mathscr{S}$ a π-system of S and consider the measurable space $(S, \sigma(\mathscr{S}))$. If $X, Y : \Omega \to S$ are two random elements such that $P(X \in B) = P(Y \in B)$ for any $B \in \mathscr{S}$, then they have the same distribution.*

The last result ensures that if the probabilities of X and Y coincide in the π-system, then they do so in the σ-algebra generated by the π-system.

Sometimes, one is interested in assuring that a particular function is a random element. In this direction, similarly as in Proposition 2.35, one can define a σ-algebra in this regard.

Proposition 2.41 *[23] Let $X : \Omega \to S$ be a function and let $(S, \mathscr{G})$ be a measurable space. Then, there exists a unique σ-algebra, $\mathscr{F}$, such that X is a random element with respect to any probability space of the form $(\Omega, \mathscr{F}, P)$ and minimal with respect to this property.*

This σ-algebra is said to be generated by X and it is denoted as $\sigma(X)$. It can be constructed by considering the σ-algebra generated by the pre-images of the subsets of $\mathscr{G}$.

Although the unit interval endowed with the Lebesgue measure is sufficient in many cases, since it allows to define sequences of random variables, sometimes

bigger probability spaces are needed. In particular, two properties that will be necessary in some specific cases will be considered. The first one allows one to have a biyection in the probability space that links any two random elements with the same distribution. This type of biyections are called measure preserving functions and can be interpreted as symmetries in the probability space.

Definition 2.42 *[186] Let $(\Omega, \mathscr{F}, P)$ be a probability space and let $\phi : \Omega \to \Omega$ be a measurable function. Then, ϕ is said to be a measure preserving function if $P(\phi^{-1}(B)) = P(B)$ for any $B \in \mathscr{F}$.*

In order to have the desirable property that will be used in Theorem 3.26, it is necessary to consider hyperfinite probability spaces. Hyperfinite probability spaces, which are particular cases of Loeb probability spaces [186], are constructed by using non-standard analysis methods [280] (although the result is a proper standard probability space [184]). In terms of cardinality, a hyperfinite probability space $(\Omega, \mathscr{F}, P)$ for which Ω is not finite is equivalent to the uncountable product of the unit interval with the Lebesgue measure (see the remark on page 5 of [184]). The reader is referred to references [184, 185, 187, 186, 280] for more information in this regard.

Theorem 2.43 *[184] Let (Ω, Σ, P) be a hyperfinite 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. The reader is referred to page 134, Proposition 9.2 in [184] for a constructive proof. Another property of interest is the saturation property, defined as follows.

Definition 2.44 *[187] A probability space (Ω, Σ, P) is said to have the saturation property if, given a pair of Polish 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 over T_1, there exists a random element $Y : \Omega \to T_2$ such that (X, Y) has distribution μ.*

The reader is referred to Section 2.1.2 for a definition of Polish spaces. Any probability space that fulfills the saturation property is called a saturated probability space. Using results in [185], it can be stated that any hyperfinite probability space is saturated [186]. In fact, this is true for any Loeb space [187].

2.3.2 Random vectors and copulas

Although different random elements will be considered along this thesis, the attention will be focused mainly on random vectors, which are finite collections of random elements whose image is contained in the real line. In particular, the basics about random vectors, their distribution, first moments, dependence modeling and sorted samples are explained.

2.3.2.1 Random vectors

As mentioned before, the most prominent example of a random element is the random variable, obtained by considering the real numbers with the Borel σ-algebra associated with the usual topology in $\mathbb{R}$. In fact, this σ-algebra, denoted by $\mathbb{B}$, can be generated by a smaller family of sets [278],

$$\mathbb{B} = \sigma\left(\{(-\infty, r) \mid r \in \mathbb{R}\}\right).$$

In many cases, working with more than one random variable is necessary. Therefore, often random vectors should be considered, which are vectors in which each component is a random variable. In this case, the Borel σ-algebra associated with the usual topology of $\mathbb{R}^n$ will be denoted as $\mathbb{B}^n$.

Definition 2.45 *[278] Let $(\Omega, \mathscr{F}, P)$ be a probability space. Then, a function $\vec{X}$: $\Omega \to \mathbb{R}^n$ is a random vector if $\vec{X}^{-1}(B) \in \mathscr{F}$ for any $B \in \mathbb{B}^n$.*

In fact, for having the measurability with respect to $(\mathbb{R}^n, \mathbb{B}^n)$, it is enough to have the measurability of each of the components with respect to $(\mathbb{R}, \mathbb{B})$. In the following, all random vectors will be considered of dimension n unless otherwise stated. The random variables associated with the components of a random vector are commonly known as their marginals. There are some properties of random vectors that are defined in terms of its marginals.

Definition 2.46 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then,*

- *$\vec{X}$ is said to be exchangeable if it has the same distribution as any permutation of its components,*
- *$\vec{X}$ and $\vec{Y}$ are in the same Fréchet class if they have the same marginal distributions.*

Any random vector, as a random element, induces a probability measure on $(\mathbb{R}^n, \mathbb{B}^n)$. In general, it is useful to work with the distribution and survival functions, defined as follows.

Definition 2.47 *[253] Let $\vec{X}$ be a random vector. Then,*

- *The distribution function of $\vec{X}$ is the function $F_{\vec{X}} : \mathbb{R}^n \to [0,1]$ defined as $F_{\vec{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$,*
- *The survival (or reliability) function of $\vec{X}$ is the function $\bar{F}_{\vec{X}} : \mathbb{R}^n \to [0,1]$ defined as $\bar{F}_{\vec{X}}(x_1, \ldots, x_n) = P(X_1 > x_1, \ldots, X_n > x_n)$.*

For simplicity, the subscript in $F_{\vec{X}}$ and $\bar{F}_{\vec{X}}$ will be omitted when the associated random vector or random variable is clear. Both the distribution and the survival function characterize the distribution of the random vector. Therefore, if two random vectors $\vec{X}$ and $\vec{Y}$ have the same distribution or survival function, then they have the same distribution. In the univariate case, distribution and survival functions are easy to characterize.

Proposition 2.48 *[278] Let $F : \mathbb{R} \to [0,1]$ and $\bar{F} : \mathbb{R} \to [0,1]$ be two functions. Then,*

- *F is the distribution function of a random variable if and only if it is increasing, right-continuous, $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$,*
- *$\bar{F}$ is the survival function of a random variable if and only if it is decreasing, left-continuous, $\lim_{x \to -\infty} F(x) = 1$ and $\lim_{x \to \infty} F(x) = 0$.*

Moreover, it is clear that, in the univariate case, $F_X(x) + \bar{F}_X(x) = 1$ holds for any random variable X and $x \in \mathbb{R}$. In addition, the quantile function can be defined for random variables.

Definition 2.49 *[278] Let X be a random variable with distribution function F. Then, its quantile function is defined as*

$$F^{-1}(p) = \inf\{x \in \mathbb{R} \mid p \leq F(x)\}.$$

When the distribution function F is strictly increasing, the quantile function F^{-1} is its inverse. Similar characterizations as in Proposition 2.48 can be done for random vectors, but they are not as simple. However, they can be derived from the latter result and Theorem 2.67.

If there exists $m \in \mathbb{R}$ such that $m - X =_{st} X - m$, then it is said that X is symmetric with respect to m. It is easy to show that in that case m is the median of X, i.e. $F^{-1}(0.5) = m$, and that the expectation of X, see the subsequent Definition 2.54, also equals m if it exists.

Also related to distribution functions, a distortion is a continuous increasing function $h : [0,1] \to [0,1]$ such that $h(0) = 0$ and $h(1) = 1$. One of the main properties of any distortion is that the composition of a distribution function and a distortion is again a distribution function. In this regard, they have been used widely in reliability theory. See, for instance, [12, 252]. A special type of distortions will be of relevance for some of the results presented in Section 5.1, since they preserve the symmetry of distributions.

Definition 2.50 *[41] A distortion h is said to be centrally symmetric if $h(t) = 1 - h(1-t)$ for any $t \in [0,1]$.*

Notice that if a distortion h is centrally symmetric and X is symmetric, $h(F_X(t))$ is the distribution function of a symmetric random variable.

The most usual classification of random vectors is between discrete and continuous. In the first case, probabilities can be computed just by doing (maybe infinite) sums, although in the second one integrals of density functions appear.

Definition 2.51 *[278] Let $\vec{X}$ be a random vector. Then, it is discrete if there exists a sequence $(\vec{x}_k, k \in \mathbb{N}) \subset \mathbb{R}^n$, such that*

$$\sum_{k=1}^{\infty} P(\vec{X} = \vec{x}_k) = 1.$$

Definition 2.52 *[278] Let $\vec{X}$ be a random vector. Then, it is continuous if its distribution function F is absolutely continuous, that is, if there exists a function $f : \mathbb{R}^n \to \mathbb{R}$ such that*

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n) d\vec{t}.$$

The function f is called the density function. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a density function of a random vector if and only if it is positive and fulfills $\int_{\mathbb{R}^n} f(\vec{x}) d\vec{x} = 1$. It must be noted that there exist random vectors that are neither discrete nor continuous.

Another concept that will be useful is the notion of support of a random vector, which is, roughly speaking, the set of values that the random vector takes that are relevant in order to compute probabilities. In fact, it is the support of the probability measure generated by the random vector.

Definition 2.53 *[262] Let $\vec{X}$ be a random vector. Then, the support (or range) of $\vec{X}$, denoted as $S(\vec{X})$, is the smallest closed set such that $P(X \in S(\vec{X})) = 1$.*

The existence of such a closed set is ensured (see Theorem 2.1 in [262]). Notice that there are other notions of support in the literature (see page 40, Remark 3 in [278]), which are based on the points in which the distribution function is increasing. However, Definition 2.53 is equivalent to considering the closure of such a notion and it is more useful for the purposes of this work.

Typically, some quantities are computed to summarize the behavior of random variables and vectors. For the location, the most usual one is the expectation (or mean), defined as follows.

Definition 2.54 *[278] Let X be a random variable. Then, its expectation, if it exists, is the value*

$$E[X] = \int_{\mathbb{R}} x dF(x).$$

The integral in the latter definition is a Riemann-Stieltjes integral, see Section 2.1.3. The condition of existence of the expectation is necessary since it is possible to have a non-convergent integral. When the random variable is discrete, then the formula simplifies to

$$E[X] = \sum_{k=1}^{\infty} x_k P(X = x_k),$$

and, if the random variable is continuous, to the expression

$$E[X] = \int_{\mathbb{R}} x f(x) dx.$$

Regarding the variability of random variables, the main related quantity is the variance, defined as the expectation of the squared difference between the variable and its expectation.

Definition 2.55 *[278] Let X be a random variable. Then, its variance, if it exists, is defined as $Var(X) = E\left[(X - E[X])^2\right]$.*

Higher variance values imply greater variability. Sometimes, the square root of the variance, known as the standard deviation, is considered. In other cases, it is of interest to study the joint variability of two random variables. In order to do that, it is usually considered the covariance.

Definition 2.56 *[278] Let X and Y be two random variables. Then, their covariance, if it exists, is defined as $Cov(X,Y) = E\left[(X - E[X])(Y - E[Y])\right]$.*

However, covariance depends not only on the dependence of the involved random variables but also increases if the variables are more disperse. In order to overcome this problem, the Pearson correlation coefficient, which normalizes the covariance by using the standard deviations of the involved random variables, is often used.

Definition 2.57 *[278] Let X and Y be two random variables. Then, their Pearson correlation coefficient, if it exists, is defined as $\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$.*

If $\rho_{X,Y} = 0$, then X and Y are said to be linearly independent. A stronger related notion is independence, defined as follows.

Definition 2.58 *[278] Let $\vec{X}_1, \ldots, \vec{X}_m$ be random vectors. Then, it is said that they are independent if $P(\vec{X}_1 \in B_1, \ldots, \vec{X}_m \in B_m) = \prod_{i=1}^{m} P(\vec{X}_i \in B_i)$ for any $B_1, \ldots, B_m \in \mathbb{B}^m$.*

Independence can also be characterized by the factorization of the distribution function as the product of the distribution functions of the marginals. Transformations of independent random vectors are independent [278]. Independence implies linear independence. In the following result, the main properties about expectation, variance, covariance and Pearson correlation coefficient are briefly exposed.

Theorem 2.59 *[278] Let X and Y be two random variables and $\lambda \in \mathbb{R}$. Then,*

- $E[\lambda] = \lambda$,
- $E[X+Y] = E[X] + E[Y]$,
- $E[\lambda X] = \lambda E[X]$,
- *If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$,*
- $Var(\lambda X) = \lambda^2 Var(X)$,
- $Cov(X,X) = Var(X)$,
- $Cov(X+\lambda, Y) = Cov(X,Y) = Cov(Y,X)$,
- $Cov(\lambda X, Y) = \lambda Cov(X,Y)$,
- $\rho_{X,Y} \in [0,1]$,
- $\rho_{X+\lambda,Y} = \rho_{\lambda X,Y} = \rho_{X,Y} = \rho_{Y,X}$,
- *If X and Y are independent, $Cov(X,Y) = \rho_{X,Y} = 0$.*

The vector that contains the expectations of the components of a random vector is called its mean vector [278]. In addition, the variability and dependence of a random vector can be summarized using a matrix known as the covariance matrix.

Definition 2.60 *[224] Let $\vec{X}$ be a random vector. Then, its covariance matrix, if it exists, is the square matrix $(\Sigma_{i,j}, i,j \in [n])$ such that $\Sigma_{i,j} = Cov(X_i, X_j)$.*

The diagonal of the covariance matrix contains the variance of the marginals. The covariance matrix of any random vector is positive semi-definite [224], see Definition 2.1.

2.3.2.2 Conditional probabilities

On some occasions, there is prior knowledge of the outcome of the random experiment associated with the probability space. In order to compute probabilities given this knowledge, the concept of conditional probability is introduced.

Definition 2.61 *[191] Let $(\Omega, \mathscr{F}, P)$ be a probability space and $A, B \in \mathscr{F}$ such that $P(B) > 0$. Then, the conditional probability of A given B, denoted as $P(A \mid B)$ is defined as*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Similarly, the conditional distribution of a random vector given B is defined as

$$F_{\vec{X} \mid B}(\vec{x}) = P(\vec{X} \leq \vec{x} \mid B) = P(\vec{X}^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) \mid B).$$

Moreover, the set B can also be associated with another random vector. In particular,

$$F_{\vec{X} \mid \vec{Y} \in B}(\vec{x}) = P(\vec{X} \leq \vec{x} \mid \vec{Y} \in B) = P(\vec{X}^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) \mid \vec{Y}^{-1}(B)).$$

If B is just a point, it will be denoted as $F_{\vec{X} \mid \vec{Y} = \vec{y}}$. For continuous random vectors, it is possible to define a conditional density function.

Definition 2.62 *[191] Let $(\vec{X}, \vec{Y})$ be a continuous random vector with density function f. Denote by $f_{\vec{Y}}$ the density function of $\vec{Y}$. Then, the conditional density of $\vec{X}$ given that $\vec{Y} = \vec{y}$, $f_{\vec{X} \mid \vec{Y} = \vec{y}}$, is defined as*

$$f_{\vec{X} \mid \vec{Y} = \vec{y}}(\vec{x}) = \frac{f(\vec{x}, \vec{y})}{f_{\vec{Y}}(\vec{y})},$$

whenever $f_{\vec{Y}}(\vec{y}) > 0$.

Using the latter conditional density, a conditional distribution function can be computed. In the general setting, where $\vec{Y}$ may be neither discrete nor continuous, it is not easy to know for which values of $\vec{y}$ the distribution function $F_{\vec{X} \mid \vec{Y}=\vec{y}}$ is well defined. At least, see Section 8.3 in [191], it is known that it is almost surely well-defined.

Theorem 2.63 *[191] Let $\vec{X}$ and $\vec{Y}$ be two random vectors of dimensions, respectively, n and m. Then, there exists $C \subseteq \mathbb{R}^m$ such that $P(\vec{Y} \in C) = 1$ and $F_{\vec{X} \mid \vec{Y}=\vec{y}}$ is well-defined for any $\vec{y} \in C$.*

In the following, all conditional distributions will be considered to be well-defined. Notice that conditional probabilities and the conditional density define a probability measure, but not a random element. However, it is common to use $[\vec{X} \mid B]$, $[\vec{X} \mid \vec{Y} \in B]$ or $[\vec{X} \mid \vec{Y} = \vec{y}]$ to denote a random vector with the considered conditional distribution. Moreover, it is possible to define a conditional expectation that is actually a random variable [191].

Definition 2.64 *[191] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Denote by $E[\vec{X} \mid \vec{y}]$ the expectation of $[\vec{X} \mid \vec{Y} = \vec{y}]$. Then, the conditional expectation of $\vec{X}$ given $\vec{Y}$ is the random variable $E[\vec{X} \mid \vec{Y}]$.*

Notice that the randomness of $E[\vec{X} \mid \vec{Y}]$ is a consequence of the randomness of $\vec{Y}$. A related notion is the mixture of distributions, which is defined as the convex linear combination of distribution functions.

Definition 2.65 *[233] Let $\vec{X}_1, \ldots, \vec{X}_m$ be random vectors. Then, $\vec{Y}$ is said to be a mixture of $\vec{X}_1, \ldots, \vec{X}_m$ if there exists a random vector $\vec{Z}$ that takes values over the permutations of the vector $(1, 0, \ldots, 0)$ and is independent of $\vec{X}_1, \ldots, \vec{X}_m$ such that*

$$Y = \sum_{i=1}^{m} Z_i \vec{X}_i.$$

It is easy to show that the distribution function of the mixture Y can be expressed as $F_{\vec{Y}}(\vec{x}) = \sum_{i=1}^{m} P(Z_i = 1) F_{\vec{X}_i}(\vec{x})$ for any $\vec{x} \in \mathbb{R}$.

2.3.2.3 Copulas

The distribution of any random vector can be decomposed into two parts, the distribution of its marginals and the dependence between them. The dependence can be expressed by means of copulas, which are distribution functions with standard uniform marginals (see Definition 2.70).

Definition 2.66 *[253] A function $C : [0,1]^n \to [0,1]$ is a copula if the following conditions are fulfilled:*

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

$$\sum_{\vec{z} \in V_{[\vec{x},\vec{y}]}} sgn(\vec{z}) C(\vec{z}) \geq 0,$$

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

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

Copulas have all the information about the dependence between the components in the associated random vector. In fact, the relation between multivariate distributions and copulas is stated by the Sklar Theorem, given below.

Theorem 2.67 *[253, 298] Let $\vec{X}$ be a random vector with marginal distributions $F_1, \ldots, F_n$ and joint distribution function F. Then, there exists a copula $C : [0,1]^n \to [0,1]$ such that*

$$F(x_1, \ldots, x_n) = C\left(F_1(x_1), \ldots, F_n(x_n)\right), \ \forall \vec{x} \in \mathbb{R}^n.$$

In addition, if $F_1, \ldots, F_n$ are continuous, the copula is unique.

Similarly, any joint survival function $\bar{F}$ can be decomposed as $\bar{F}(x_1,\ldots,x_n) = \bar{C}(\bar{F}_1(x_1),\ldots,\bar{F}_n(x_n))$ for any $\vec{x} \in \mathbb{R}^n$, where $\bar{F}_1,\ldots,\bar{F}_n$ and $\bar{C}$ is a copula, usually called the survival copula of the random vector.

If the marginal distributions $F_1,\ldots,F_n$ are not continuous, then there exists a unique subcopula C' defined over $F_1(\mathbb{R}) \times \cdots \times F_n(\mathbb{R})$ such that $F(x_1,\ldots,x_n) = C'(F_1(x_1),\ldots,F_n(x_n))$ for any $\vec{x} \in \mathbb{R}$. The reader is referred to [105] for more information in this regard.

Any copula is a continuous function [106, 118, 253]. Prominent examples of copulas are the product copula $C_\prod(\vec{x}) = \prod_{i=1}^n x_i$, which leads to independence, and the minimum copula $C_{\min}(\vec{x}) = \min(\vec{x})$, which is associated with comonotone (perfectly positive dependent) random variables. In the bivariate case, the case of the Łukasiewicz t-norm $C_L(x_1,x_2) = \max(0, x_1 + x_2 - 1)$ is also relevant, which is associated with countermonotone (perfectly negative dependent) random variables. In addition, any copula fulfills

$$\max\left(0, 1-n+\sum_{i=1}^{n} x_i\right) \leq C(\vec{x}) \leq \min(\vec{x}),$$

which is known as the Fréchet Hoeffding bounds [118, 253].

When the copula of Theorem 2.67 can be chosen to be the same for two random vectors, it is said that they have the same copula, which means that they share the same dependence structure. It is worth noting that if two random vectors have the same marginals distributions and copula, they also have the same distribution. Furthermore, it is important to mention that an increasing transformation of the marginals of a random vector does not change its copula [253].

If the copula is absolutely continuous, then one can consider the density copula $c : [0,1]^n \to \mathbb{R}$ as the function such that $C(\vec{x}) = \int_{\vec{0} \leq \vec{t} \leq \vec{x}} c(\vec{t}) d\vec{t}$.

Example 2.68 *[118] Consider the following examples of copulas.*

- *A copula C is said to be a Farlie-Gumbel-Morgenstern copula if $C(x,y) = xy + \lambda x(1-x)y(1-y)$ for any $x,y \in [0,1]$, where $\lambda \in [-1,1]$,*

- *A copula C is said to be a Gaussian copula if its density copula has the ex-*

pression

$$c(x_1,\dots,x_n) = \frac{1}{\sqrt{|R|}} exp\left(-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(x_1)\\ \dots \\ \Phi^{-1}(x_n)\end{pmatrix}^t (R^{-1}-I)\begin{pmatrix}\Phi^{-1}(x_1)\\ \dots \\ \Phi^{-1}(x_n)\end{pmatrix}\right),$$

where R is a positive definite matrix, I is the identity matrix and Φ^{-1} *is the quantile function of a standard Gaussian distribution,*

- *A copula C is said to be a T-copula if its density copula has the expression:*

$$c(x_1,\dots,x_n) = \frac{f_{\nu,R}(t_\nu^{-1}(x_1),\dots,t_\nu^{-1}(x_n))}{\prod_{i=1}^n f_\nu(t_\nu^{-1}(x_i))},$$

where R is a positive definite matrix with constant diagonal equal to 1, t^{-1} *is the quantile function of a standard Student's t-distribution,* f_ν *is the density function of a standard Student's t-distribution and* $f_{\nu,R}$ *is the distribution function of a multivariate Student's t-distribution with mean vector* $\vec{0}$ *and dispersion matrix R.*

2.3.2.4 Relevant families of probability distributions

This section is devoted to briefly introduce the distributions of random variables and random vectors that are going to be considered in some results and examples, which can be found in [176, 177, 201]. Bernoulli distribution is one of the most simple cases.

Definition 2.69 *A random variable X is said to have Bernoulli distribution if* $P(X = 0) = 1 - p$ *and* $P(X = 1) = p$ *with* $p \in [0,1]$.

Moving to continuous distributions, one of the most relevant cases is the uniform distribution. Given a set, a (continuous) uniform distribution is an absolutely continuous probability distribution such that its density function is constant on the set and 0 everywhere else. Recall that $\lambda(A)$ denotes the Lebesgue measure on $\mathbb{R}^n$ of $A \subseteq \mathbb{R}^n$. Formally, the following definition will be considered.

Definition 2.70 *Given a bounded and Lebesgue-measurable subset $A \subset \mathbb{R}^n$ with Lebesgue measure $\lambda(A) > 0$, a uniform distribution over A is defined as the probability measure P:*

$$P(B) = \frac{\lambda(B \cap A)}{\lambda(A)},$$

for each Borel-measurable subset $B \subseteq \mathbb{R}^n$.

The particular case of A being the unit interval $[0,1]$ is known as the standard uniform distribution. Sums of independent standard uniform random variables follow the Irwin-Hall distribution.

Definition 2.71 *Let X be a random variable with density function*

$$f(x) = \frac{1}{(m-1)!} \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \frac{m!}{k!(m-k)!} (x-k)^{m-1},$$

if $x \in [0,m]$ and $f(x) = 0$ elsewhere, where $\lfloor x \rfloor$ denotes the integer part of x and $m \in \mathbb{N}$. Then, X is said to have Irwin-Hall distribution.

Proposition 2.72 *[225] Let $X_1, \ldots, X_m$ be m independent random variables with standard uniform distribution. Then, $\sum_{i=1}^{m} X_i$ has Irwin-Hall distribution with parameter m.*

In addition, the order statistics of independent standard uniform random variables (see subsequent Section 2.3.2.5), follow Beta distributions.

Definition 2.73 *Let X be a random variable with density function*

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},$$

if $x \in [0,1]$ and $f(x) = 0$ elsewhere, with $\alpha, \beta \in \mathbb{R}^+$, $B(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$. Then, X is said to have Beta distribution and it is denoted as $X \sim B(\alpha,\beta)$.

For random vectors, the most prominent distribution is the multivariate Gaussian distribution. A continuous random vector $\vec{X}$ has a multivariate Gaussian distribution if any linear combination of its components has univariate Gaussian distribution [224]. The distribution of this type of vectors is characterized by its mean vector $\vec{\mu}$ and its covariance matrix Σ.

Definition 2.74 *Let $\vec{X}$ be a random vector with density function*

$$f(\vec{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{(\vec{x}-\vec{\mu})^t \Sigma^{-1} (\vec{x}-\vec{\mu})}{2}\right),$$

for any $\vec{x} \in \mathbb{R}$ where $\vec{\mu} \in \mathbb{R}^n$ and Σ is a positive semi-definite matrix. Then, $\vec{X}$ is said to have multivariate Gaussian distribution and it is denoted as $\vec{X} \sim N(\vec{\mu}, \Sigma)$.

For dimension 1, Gaussian random variables with null mean and variance equal to one are called standard Gaussian random variables. A generalization of Gaussian random variables can be done by considering a shape parameter, as the generalized Gaussian distribution does.

Definition 2.75 *[120] Let X be a random variable with density function*

$$f(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left(\frac{|x-\mu|}{\alpha}\right)\right),$$

for any $x \in \mathbb{R}$ with $\mu \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}^+$. Then, X is said to have generalized Gaussian distribution.

Another distribution related to the standard Gaussian distribution, which appears in some hypothesis tests and is equivalent to the first one when its parameter ν goes to infinite, is the Student's t-distribution.

Definition 2.76 *Let X be a random variable with density function*

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi r}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

for any $x \in \mathbb{R}$ with $\nu \in \mathbb{R}^+$. Then, X is said to have Student's t-distribution with ν degrees of freedom.

Similarly, the Chi-squared distribution is also related to the standard Gaussian distribution. In particular, the sum of the square of independent Gaussian random variables has this type of distribution.

Definition 2.77 *Let X be a random variable with density function*

$$f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}},$$

for any $x \in \mathbb{R}$ with $k \in \mathbb{N}$. Then, X is said to have Chi-squared distribution with k degrees of freedom.

The Chi-squared distribution is a particular case of a more general distribution, the Gamma distribution.

Definition 2.78 *Let X be a random variable with density function*

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{\beta x},$$

for any $x \in \mathbb{R}$ with $\alpha, \beta \in \mathbb{R}^+$ Then, X is said to have Gamma distribution with shape parameter α and rate parameter β and it is denoted as $X \sim \Gamma(\alpha, \beta)$.

For random variables taking values over the positive real numbers, one of the most well-known distributions is the exponential distribution.

Definition 2.79 *Let X be a random variable with density function*

$$f(x) = \lambda e^{-\lambda x},$$

for any $x \in \mathbb{R}^+$ and $f(x) = 0$ elsewhere with $\lambda \in \mathbb{R}^+$. Then, X is said to have exponential distribution with parameter λ and it is denoted as $X \sim Exp(\lambda)$.

The Laplace distribution is a version of the exponential distribution that also considers negative values.

Definition 2.80 *Let X be a random variable with density function*

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right),$$

for any $x \in \mathbb{R}$ with $\mu \in \mathbb{R}$ and $b \in \mathbb{R}^+$. Then, X is said to have Laplace distribution with location parameter μ and scale parameter b.

Finally, some distributions are defined in terms of trigonometrical functions.

Definition 2.81 *Let X be a random variable with density function*

$$f(x) = \frac{1}{4s}\left(sech\left(\frac{x-\mu}{2s}\right)\right)^2,$$

for any $x \in \mathbb{R}$ with $\mu \in \mathbb{R}$, $b \in \mathbb{R}^+$ and sech denoting the hyperbolic secant function. Then, X is said to have logistic distribution.

Definition 2.82 *[115] Let X be a random variable with density function*

$$f(x) = \frac{1}{2}sech\left(\frac{\pi}{2}x\right),$$

for any $x \in \mathbb{R}$ with sech denoting the hyperbolic secant function. Then, X is said to have hyperbolic secant distribution.

2.3.2.5 Order statistics

In the literature, many methods involve the ordering, from the smallest to the greatest, of the values of independent and identically distributed random variables $X_1, \ldots, X_n$. These are known as order statistics and are denoted as $X_{(1)}, \ldots, X_{(n)}$, where $X_{(1)} = \min(X_1, \ldots, X_n)$ and $X_{(n)} = \max(X_1, \ldots, X_n)$.

The distribution functions of the order statistics are easy to compute, since they have a simple expression in terms of the distribution function of the marginals.

Proposition 2.83 *[104] Let $X_1, \ldots, X_n$ be independent and identically distributed random variables. Then, the distribution function of the order statistic $X_{(k)}$, denoted as $F_{(k)}$, is*

$$F_{(k)}(x) = \sum_{i=k}^{n} \binom{n}{i} F(x)^i (1-F(x))^{n-i} = \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} F(x)^i (1-F(x))^{n-i},$$

where $F(t)$ denotes the distribution function of X_1.

If the distribution is continuous, then the joint density of all order statistics is $n!f(x)^n$ on the set $\{\vec{x} \in \mathbb{R}^n \mid x_1 \leq \cdots \leq x_n\}$ [104].

The most studied distribution in the field of order statistics is the uniform distribution, generally over the unit interval, since the order statistics have simple distributions and some quantities, such as linear combinations of them, can also be easily characterized. In particular, the order statistics follow Beta distributions.

Proposition 2.84 *[104] Let $U_1,\dots,U_n$ be a sequence of independent random variables with standard uniform distribution. Then, $U_{(k)}$ has Beta distribution with parameters $\alpha = k$ and $\beta = n+1-k$.*

In addition, more involved results about linear combinations of the order statistics of uniform random variables can be proved. In the simplest case, the linear coefficients are strictly positive. The following is Equation (6.5.5) in [104].

Proposition 2.85 *[104] Let $U_{(1)},\dots,U_{(n)}$ be the order statistics of a sequence of n standard uniform random variables and $a_1,\dots,a_n > 0$. Then, the distribution function of $T = \sum_{i=1}^n a_i U_{(i)}$ has the following expression,*

$$F_T(x) = 1 - \sum_{i=1}^{m} \frac{(c_i - x)^n}{c_i \prod_{j\neq i}(c_i - c_j)} \mathbb{I}_{x\in[0,c_1]},$$

where $c_i = \sum_{k=i}^n a_k$, $\mathbb{I}$ denotes the indicator function and m is the largest integer such that $x \leq c_m$.

Unfortunately, if any of the coefficients of the linear combination is 0, the expression is more complicated.

Proposition 2.86 *[104] Let $U_{(1)},\dots,U_{(n)}$ be the order statistics of a sample of n standard uniform random variables and $a_1,\dots,a_n \geq 0$. Denote as $a_{n_1},\dots,a_{n_k}$ the coefficients such that $a_{n_i} > 0$ with $i \in [k]$. Then, the distribution function of $T = \sum_{i=1}^n a_i U_{(i)}$ have the following expression,*

$$F_T(x) = 1 - \sum_{i=1}^{m} \frac{g_i^{(n_1 - n_{i-1} - 1)}(c_{n_i}, x)}{(n_i - n_{i-1} - 1)!} \mathbb{I}_{x\in[0,c_1]},$$

where $c_i = \sum_{k=i}^n a_k$, $\mathbb{I}$ denotes the indicator function, m is the largest integer such that $x \leq c_m$ and $g_i^{(s)}(c,x)$ denotes

$$g_i^{(s)}(c,x) = \frac{\partial^s}{\partial c^s} \frac{(c-x)^n}{c\prod_{j\neq i}(c - c_j)}.$$

However, in many situations, the distribution is not uniform. In these cases, many asymptotic results have been developed. In the following, a result that combines the results of [306], which gives a convenient definition of the involved sequence, and [307], in which the uniform convergence part is proved, is provided.

Theorem 2.87 *[306, 307] Let $X_1,\ldots,X_n$ be a sequence of independent and identically distributed random variables with density function f and cumulative distribution function F such that f is continuous and strictly positive in $F^{-1}(0,1)$ and there exists $\varepsilon > 0$ for which $\lim_{x\to\infty} |x|^{\varepsilon}\left[1-F(x)+F(-x)\right] = 0$. Then, for any $\delta > 0$,*

$$\lim_{n\to\infty}(n+2)Cov\left(X_{((n+1)q)},X_{((n+1)p)}\right) = \frac{(1-p)q}{f\left(F^{-1}(p)\right)f\left(F^{-1}(q)\right)},$$

uniformly for $p,q \in [\delta,1-\delta]$ such that $p \leq q$.

As a consequence, the inverse of the covariance matrix of the order statistics, when n goes to infinity, can be approximated as $\Sigma^{-1} \sim (n+1)(n+2)DQD$ [306], where D is a diagonal matrix satisfying that $D_{i,i} = f\left(F^{-1}(\frac{i}{n+1})\right)$ for any $i \in [n]$ and Q is the matrix:

$$Q = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \ldots \\ -1 & 2 & -1 & 0 & 0 & \ldots \\ 0 & -1 & 2 & -1 & 0 & \ldots \\ 0 & 0 & -1 & 2 & -1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{2.1}$$

The inverse matrix of Q has a simple structure. In particular, $(Q^{-1})_{i,j} = \frac{i(n+1-j)}{n+1}$ for any $i,j \in [n]$ (see [173]).

2.3.3 Stochastic processes

Random vectors can be seen as finite collections of random variables. In some contexts, it is necessary to consider sets of random variables, defined in the same probability space and indexed over a particular set, that can have infinite cardinality. These sets of random variables are known as stochastic processes.

Definition 2.88 *[168, 213, 261] Let T be a set. A stochastic order process T is a collection of random variables $(X_t, t \in T)$ defined in the same probability space.*

The index set is denoted by T since it is usually referred to time. Typically, it is chosen to be $\mathbb{R}^+$ or $\mathbb{N}$, representing, respectively, continuous or discrete time. However, the set T can be, in fact, any set such as a graph [282] or the real plane [324].

For each $\omega \in \Omega$, $(X_t(\omega), t \in T)$ can be seen as a function from T to the real numbers. Each of such values is called a sample path of the stochastic order, which is similar to the concept of observation of a random variable or random vector. Sometimes, the properties of sample paths, such as continuity, are especially relevant.

Notice that, if a finite subset of the index set $T_0 \subseteq T$ is considered, the collection $(X_t, t \in T_0)$ can be seen as a random vector. The distributions associated with these random vectors are known as the finite-dimensional distributions of the stochastic process. In addition, any consistent collection of finite-dimensional distributions characterizes the distribution of a stochastic process by the Kolmogorov Extension Theorem [194].

When the stochastic process is indexed by time, both in the cases $T = \mathbb{N}$ and $T = \mathbb{R}^+$, many properties can be defined.

Definition 2.89 *[153, 261] Let $(X_t, t \in \mathbb{N})$ be a stochastic process. Then,*

- *If $P(X_{t_n} \in B_n \mid X_{t_1} \in B_1, \ldots, X_{t_{n-1}} \in B_{n-1}) = P(X_{t_n} \in B_n \mid X_{t_{n-1}} \in B_{n-1})$ for any $n \in \mathbb{N}$, $B_1, \ldots, B_n \in \mathbb{B}$ and $t_1, \ldots, t_n \in \mathbb{N}$ such that $t_1 \leq \cdots \leq t_n$ for which such a conditional probability exists, then $(X_t, t \in \mathbb{N})$ is said to fulfill the Markov property,*

- *If $(X_{t_1}, \ldots, X_{t_n}) =_{st} (X_{t_1+k}, \ldots, X_{t_n+k})$ for any $t_1, \ldots, t_n, n, k \in \mathbb{N}$, then $(X_t, t \in \mathbb{N})$ is said to be stationary,*

- *If $E[X_t] = E[X_1]$ for any $t \in \mathbb{N}$ and*

$$\frac{1}{n} \sum_{t=0}^{n} X_t \rightarrow_{a.s} E[X_1],$$

then $(X_t, t \in \mathbb{N})$ is said to be (mean) ergodic,

- *If* $E[X_n \mid X_1,\ldots,X_{n-1}] = (\leq,\geq)X_{n-1}$, *then* $(X_t, t \in \mathbb{N})$ *is said to be a martingale (supermartingale, submartingale).*

When considering a continuous time, similar notions can be introduced.

Definition 2.90 *[153, 168, 261] Let* $(X_t, t \in \mathbb{R}^+)$ *be a stochastic process. Then,*

- *If* $P(X_{t_n} \in B_n \mid X_{t_1} \in B_1,\ldots,X_{t_{n-1}} \in B_{n-1}) = P(X_{t_n} \in B_n \mid X_{t_{n-1}} \in B_{n-1})$ *for any* $n \in \mathbb{N}$, $B_1,\ldots,B_n \in \mathbb{B}$ *and* $t_1,\ldots,t_n \in \mathbb{R}^+$ *such that* $t_1 \leq \cdots \leq t_n$ *for which such a conditional probability exists, then* $(X_t, t \in \mathbb{R}^+)$ *is said to fulfill the Markov property,*

- *If* $(X_{t_1},\ldots,X_{t_n}) =_{st} (X_{t_1+k},\ldots,X_{t_n+k})$ *for any* $n \in \mathbb{N}$ *and* $t_1,\ldots,t_n,k \in \mathbb{R}^+$, *then* $(X_t, t \in \mathbb{R}^+)$ *is said to be stationary,*

- *If* $E[X_t] = E[X_0]$ *for any* $t \in \mathbb{R}^+$ *and*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T X_t dt =_{a.s} E[X_0],$$

  then $(X_t, t \in \mathbb{R}^+)$ *is said to be (mean) ergodic,*

- *If* $E[X_{t_n} \mid X_{t_1},\ldots,X_{t_{n-1}}] = (\leq,\geq)X_{t_{n-1}}$, *for any* $t_1,\ldots,t_n \in \mathbb{R}^+$ *such that* $t_1 \leq \cdots \leq t_n$, *then* $(X_t, t \in \mathbb{R}^+)$ *is said to be a martingale (supermartingale, submartingale),*

- $(X_t, t \in \mathbb{R}^+)$ *is said to fulfill the reflection principle if* $P\left(\sup_{0\leq t\leq s} X_t \geq a\right) = 2P(X_s \geq a)$ *for any* $a, s \in \mathbb{R}^+$.

These properties have clear and relevant interpretations regarding the behavior of the stochastic process.

Markov property states that, given some values of the stochastic process, the future values depend only on the last known value. This is usually interpreted as the past and future being conditionally independent given the value of the present. The most remarkable examples are Markov chains, both in discrete and continuous time [255], or other processes such as the Brownian motion or the Ornstein-Uhlenbeck process [211]. In the case of discrete Markov chains, when $T = \mathbb{N}$, a

value $\lambda \in \mathbb{R}$ is said to be an absorbent state if $P(X_n = \lambda \mid X_k = \lambda) = 1$ for any $n,k \in \mathbb{N}$ such that $k < n$.

Stationary stochastic processes are temporary homogeneous, thus their distribution does not change with time. This is a relevant property in time series analysis, since it allows to apply methods that have been fitted using past data to new values, since one expects a similar behavior. Some examples of stationary processes are the Ornstein-Uhlenbeck process or some types of ARMA models [156].

Ergodicity is related to the estimation of a constant mean in the process. Of course, the mean can be estimated when having independent observations of the process, but in many applied cases this is not possible. However, in an ergodic process the mean can be estimated just by considering one observation and computing the associated sum or integral. For instance, any stationary process $(X_t, t \in \mathbb{R}^+)$ for which $\mathrm{Cov}(X_0, X_h)$ goes to 0 fast enough when t goes to infinity is ergodic (see condition [3.1.15] in [156]). On the other hand, the Brownian motion is not ergodic. Finally, a process is a martingale if, given some values of the process, the expectation of the future only depends on the latest known value and is equal to that value. Brownian motion is a martingale, while the Ornstein-Uhlenbeck process is not.

2.3.4 Random sets

Random sets are a generalization of random vectors in which the values they take are subsets rather than elements of the considered space. These type of random structures can appear when dealing with imprecise observations [236] or with sets defined by stochastic processes [144]. They have applications in different areas [53, 240, 336]

The most usual ones are the so-called closed random sets over $\mathbb{R}^n$, considering as the set of values

$$F^n = \{C \subseteq \mathbb{R}^n \mid \text{C is closed}\}.$$

Along this thesis, all random sets will be considered to be closed, non-empty and Effros measurable.

Definition 2.91 *[239] Given a probability space $(\Omega, \mathscr{F}, P)$, a function $S_X : \Omega \to F^n$*

is a Effros measurable random set if

$$\{\omega \in \Omega \mid S_X(\omega) \cap O \neq \emptyset\} \in \mathscr{F},$$

for any open set $O \subseteq \mathbb{R}^n$.

The distribution of any random set can be determined by the capacity functional, which is the probability of intersecting compact sets. As a consequence of the Choquet Theorem, see [254], the following result holds.

Theorem 2.92 *Let* S_X *and* S_Y *be two random sets of dimension n. Then, if* $P(S_X \cap K \neq \emptyset) = P(S_Y \cap K \neq \emptyset)$ *for any compact* $K \subseteq \mathbb{R}^n$, S_X *and* S_Y *have the same distribution.*

For any random set, it is possible to compute the probability of intersecting or being contained in any open or closed set. In addition, the infimum and supremum of random sets are random variables. The following is a particular case of Theorem 2.27 in [239].

Theorem 2.93 *Let S be a bounded random set. Then,* $\inf_{\vec{x} \in \mathbb{R}^n} S$ *is a random variable.*

A classical problem in random set theory is to study the selections of a given random set. A selection of S_X is a random vector $\vec{X}$ such that $P(\vec{X} \in S_X) = 1$ [254]. Intuitively, it is the result of choosing a random element that is (almost surely) contained in the random set. The most easy way to obtain a selection is by applying a selection operator, defined as follows.

Definition 2.94 *[239] A function* $f : F^n \to \mathbb{R}^n$ *is said to be a selection operator if it is measurable and* $f(C) \in C$ *for every* $C \in F^n$.

Obviously, $f(S_X)$ is always a selection of the random set S_X. It is known that there exists always a sequence of selection operators such that the closure of their composition with a random set equals the random set.

Proposition 2.95 *[239] There exists a sequence of selection operators* $\{f_i, i \in \mathbb{N}\}$ *such that* $C = cl\{f_i(C), i \in \mathbb{N}\}$ *for any* $C \in F^n$.

Of course, one has, under the conditions of the latter result, that $S_X = cl\{f_i(S_X), i \in \mathbb{N}\}$. Recall that the closure of a set C is defined as the limit points of C, where $\vec{x}$ is a limit point of C if it can be expressed as $\vec{x} = \lim_{j\to\infty} \vec{x}_j$ with $\{\vec{x}_j, j \in \mathbb{N}\} \subseteq C$.

The notion of expectation, well established for random vectors, is not straightforward for random sets. One of the most used ones is the selection expectation, defined by considering integrable selections. Consider the p-norm for random vectors defined as $E\left[||\vec{X}||_2^p\right]$, where $||\cdot||_2$ denotes the Euclidean norm.

Definition 2.96 *[239] Let S_X be a random set and $\vec{X}$ a selection of S_X. $\vec{X}$ is said to be a p-integrable selection of S_X if $E\left[||\vec{X}||_2^p\right]$ is finite.*

A random set S_X is p-integrable if there exists at least one p-integrable selection. For simplicity, 1-integrability will simply be denoted as integrability. Any p-integrable random set can be expressed as the closure of p-integrable selections as follows.

Proposition 2.97 *[239] Let S_X be a p-integrable random set. Then, there exists a sequence of p-integrable selections $(\vec{X}_i, i \in \mathbb{N})$ such that $S_X = cl\{\vec{X}_i, i \in \mathbb{N}\}$.*

Moreover, any p-integrable selection of S_X can be approximated by a sequence that is a modification of the latter one.

Theorem 2.98 *[239] Let S_X be a random set and $(\vec{X}_i, i \in \mathbb{N})$ a sequence of p-integrable selections such that $S_X = cl\{\vec{X}_i, i \in \mathbb{N}\}$. Then, for any p-integrable selection $\vec{X}$ of S_X and any $\varepsilon > 0$, there exists a finite measurable partition $A_1, \ldots, A_m$ of Ω such that:*

$$E\left[\vec{X} - \sum_{i=1}^{m} \mathbb{I}_{A_i}\vec{X}_i\right] < \varepsilon,$$

where $\mathbb{I}$ denotes the indicator function.

It is important to remark that, for $p = 2$, the latter result implies the convergence in L^2 of the sequence. This, in turn, implies convergence in distribution [222], see the paragraph right before Proposition 2.107. Finally, the selection expectation can be introduced.

Definition 2.99 *[254] Let S_X be an integrable random set. Then, the selection expectation is defined as:*

$$E[S_X] = cl\{E[\vec{X}] \mid \vec{X} \text{ integrable selection of } S_X\}.$$

2.3.5 Stochastic orders

Stochastic orders are relations between probability distributions that aim to order them, typically, in terms of location, variability or dependence. They can be seen as stronger versions of inequalities regarding the expectation, the variance or the Pearson correlation coefficient, respectively.

Since any random element induces a probability distribution, stochastic orders are usually used to order random elements. However, there are some of them, such as Statistical Preference, that order random variables but not distributions.

Actually, it is not easy to find a proper definition of a general stochastic order in the literature, even in the main reference book on the topic [295]. In order to clarify the concept, the following definition will be used.

Definition 2.100 *Let (Ω, Σ, P) be a probability space and let (S, F) be a measurable space. Denote by L the set of all measurable functions from Ω to S. A binary relation $\leq_{so} \subseteq L \times L$ is a stochastic order if:*

- *For any $f, g, h \in L$ such that $f \leq_{so} g$ and $g \leq_{so} h$, it holds $f \leq_{so} h$,*
- *For any $f, g \in L$, $f \leq_{so} g$ and $g \leq_{so} f$ holds if and only if f and g have the same distribution.*

That is, the notion of stochastic order is similar to the one of partial order (see Section 2.2.5.2) but considers classes of equivalence with respect to having the same distribution for the reflexivity and the antisymmetry.

Although the latter definition is stated in terms of random elements, most stochastic orders are defined for random vectors or random variables. In the subsequent sections, the main three types of stochastic orders (location, variability and dependence) are introduced.

2.3.5.1 Location stochastic orders

Typical stochastic orders consider the location of the variables as the criterion for the order. In fact, the most used stochastic order, the usual stochastic order, is defined in terms of expectations of increasing functions.

Definition 2.101 *Let $\vec{X}$ and $\vec{Y}$ be two random vectors. If $E[\varphi(\vec{X})] \leq E[\varphi(\vec{Y})]$ for any measurable increasing function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $E[\varphi(\vec{X})]$ and $E[\varphi(\vec{Y})]$ exist, then $\vec{X}$ is said to be smaller than or equal to $\vec{Y}$ in the usual stochastic order and it is denoted as $\vec{X} \leq_{st} \vec{Y}$.*

For random variables, the usual stochastic order is also known as Stochastic Dominance and it is equivalent to the pointwise ordering of the distribution function, that is, $X \leq_{st} Y \iff F_X(x) \geq F_Y(x) \forall x \in \mathbb{R}$ [295].

A simple way to construct two random vectors ordered by means of the usual stochastic order is to consider $\vec{X}$ and $\vec{Y}$ such that $P(\vec{X} \leq \vec{Y}) = 1$. If this is the case, it is said that $\vec{X}$ is almost surely smaller than or equal to $\vec{Y}$ and it is denoted as $\vec{X} \leq_{a.s.} \vec{Y}$. Trivially, $\vec{X} \leq_{a.s.} \vec{Y} \implies \vec{X} \leq_{st} \vec{Y}$.

In addition to the comparison of expectations of increasing functions, there are two other main equivalent definitions for the usual stochastic order, given in the following result.

Theorem 2.102 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then, the following statements are equivalent.*

1. $\vec{X} \leq_{st} \vec{Y}$,
2. $P(\vec{X} \in U) \leq P(\vec{Y} \in U)$ *for any upper set* $U \subseteq \mathbb{R}^n$,
3. *There exist two random vectors defined in the same probability space $\hat{\vec{X}}$ and $\hat{\vec{Y}}$ such that $\hat{\vec{X}} =_{st} \vec{X}$, $\hat{\vec{Y}} =_{st} \vec{Y}$ and $\hat{\vec{X}} \leq_{a.s.} \hat{\vec{Y}}$.*

If the usual stochastic order holds between two random vectors, it also holds for the associated marginals [295]. That is, if $\vec{X} \leq_{st} \vec{Y}$, then $X_i \leq_{st} Y_i$ for any $i \in [n]$. The other implication is not true in general, but it holds when having the same copula.

Proposition 2.103 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors with the same copula such that $X_i \leq_{st} Y_i$ for any $i \in [n]$. Then, $\vec{X} \leq_{st} \vec{Y}$.*

The usual stochastic order has good properties with respect to conditional distributions. In particular, if the conditional distributions are ordered for each possible value of another random vector, then the non-conditional distributions are also ordered.

Proposition 2.104 *[295] Let $\vec{X}$, $\vec{Y}$ and $\vec{Z}$ be three random vectors. If $[\vec{X} \mid \vec{Z} = \vec{z}] \leq_{st} [\vec{Y} \mid \vec{Z} = \vec{z}]$ for any $\vec{z} \in S(\vec{Z})$, then $\vec{X} \leq_{st} \vec{Y}$.*

A more involved result can be proved by considering random variables depending on a parameter.

Theorem 2.105 *[295] Let $X(\theta)$ be a random variable depending on a parameter. Let Θ_X and Θ_Y be two random variables with the same support. Then, if $X(\theta_1) \leq_{st} X(\theta_2)$ whenever $\theta_1 \leq \theta_2$ and $\Theta_X \leq_{st} \Theta_Y$, it holds that $X(\Theta_X) \leq_{st} X(\Theta_Y)$.*

Another relevant property is that the usual stochastic order is closed under the application of an increasing function.

Proposition 2.106 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors such that $\vec{X} \leq_{st} \vec{Y}$ and consider a measurable increasing function $f : \mathbb{R}^n \to \mathbb{R}^m$. Then, $f(\vec{X}) \leq_{st} f(\vec{Y})$.*

Finally, the usual stochastic order is preserved when considering sequences of random vectors. In the following $\vec{X}_n \to_d \vec{X}$ will denote convergence in distribution ($\lim_{n\to\infty} F_{\vec{X}_n}(\vec{t}) = F_{\vec{X}}(\vec{t})$ for any $\vec{t} \in \mathbb{R}^n$), $X_n \to_{a.s.} X$ will denote almost surely convergence ($P(\lim_{n\to\infty} \vec{X}_n = \vec{X}) = 1$) and $X_n \to_{L^2} X$ will denote convergence in L^2 ($\lim_{n\to\infty} E[(X_n - X)^2] = 0$) [222, 278].

Proposition 2.107 *[295] Let $(\vec{X}_n, n \in \mathbb{N})$ and $(\vec{Y}_n, n \in \mathbb{N})$ be two sequences of random vectors such that $\vec{X}_n \leq_{st} \vec{Y}_n$ for any $n \in \mathbb{N}$. If $\vec{X}_n \to_d \vec{X}$ and $\vec{Y}_n \to_d \vec{Y}$, then $\vec{X} \leq_{st} \vec{Y}$.*

Since almost surely convergence and convergence in L^2 imply convergence in distribution [278], latter result also holds for those cases. Obviously, there exist cases where the random quantities cannot be ordered in the usual stochastic order. In order to avoid such limitations, weaker stochastic orders can be used. The most common alternative to the usual stochastic order is the increasing convex stochastic order, defined as follows.

Definition 2.108 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. If $E[\varphi(\vec{X})] \leq E[\varphi(\vec{Y})]$ for any measurable increasing convex (symmetric increasing convex) function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $E[\varphi(\vec{X})]$ and $E[\varphi(\vec{Y})]$ exist, then $\vec{X}$ is said to be smaller than or equal to $\vec{Y}$ in the increasing convex (symmetric increasing convex) stochastic order and it is denoted as $\vec{X} \leq_{icx} (\leq_{sym-icx}) \vec{Y}$.*

Throughout this thesis, the increasing convex stochastic order will be referred to as simply the increasing convex order. Trivially, the usual stochastic order implies the increasing convex order. In addition, it is also clear that $\vec{X} \leq_{icx} \vec{Y} \implies E[\vec{X}] \leq E[\vec{Y}]$. Similarly to the case of the usual stochastic order, there are some equivalent conditions for the increasing convex order.

Theorem 2.109 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then, the following statements are equivalent.*

1. $\vec{X} \leq_{icx} \vec{Y}$,
2. *There exist two random vectors $\hat{\vec{X}}$ and $\hat{\vec{Y}}$ defined in the same probability space holding $\hat{\vec{X}} =_{st} \vec{X}$, $\hat{\vec{Y}} =_{st} \vec{Y}$ and $E\left[\hat{\vec{Y}} \,\middle|\, \hat{\vec{X}}\right] \geq \hat{\vec{X}}$.*

Moreover, if $\vec{X}$ and $\vec{Y}$ are random variables (denoted as X and Y), then the following statement is also equivalent.

3. $\int_x^\infty \bar{F}_X(t)dt \leq \int_x^\infty \bar{F}_Y(t)dt$, $\forall x \in \mathbb{R}$.

On the other hand, sometimes it is necessary to have location stochastic orders that are stronger than the usual stochastic order. One of the cases that is relevant in reliability theory [256], is the cumulative hazard order, denoted by $\vec{X} \leq_{ch} \vec{Y}$. Its definition, which is restricted to positive random vectors, is quite complicated and

involves integrals of failure rates. The reader is referred to Section 6.C in [295] to more information in this regard. The main result of interest states that this order is stronger than the usual stochastic order in the multivariate case and equivalent in the univariate one.

Proposition 2.110 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then, $\vec{X} \leq_{ch} \vec{Y} \implies \vec{X} \leq_{st} \vec{Y}$. Moreover, if X and Y are two random variables, $X \leq_{ch} Y \iff X \leq_{st} Y$.*

There are also many other notions of location stochastic orders. In the next definition, some of them are briefly introduced.

Definition 2.111 *Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then,*

- *[199] If $X_i \geq_{st} Y_i$ for any $i \in [n]$, it is said that $\vec{X}$ is smaller than or equal to $\vec{Y}$ in the componentwise stochastic order and it is denoted as $\vec{X} \geq_{cst} \vec{Y}$,*
- *[199] If for any $\lambda_1, \ldots, \lambda_n \in \mathbb{R}^+$, $\sum_{i=1}^n \lambda_i X_i \leq_{st} \sum_{i=1}^n \lambda_i Y_i$, it is said that $\vec{Y}$ has Linear Stochastic Dominance over $\vec{X}$ and it is denoted by $\vec{X} <_{LSD} \vec{Y}$,*
- *[279] When $\vec{X}$ and $\vec{Y}$ are defined in the same probability space, if $P(\vec{X} < \vec{Y}) + \frac{1}{2}P(\vec{X} = \vec{Y}) > 0.5$ (componentwise), it is said that $\vec{Y}$ is statistically preferred to $\vec{X}$ and it is denoted as $\vec{X} \leq_{SP} \vec{Y}$,*
- *[295] If $P(\vec{X} = \vec{t})P(\vec{Y} = \vec{s}) \leq P(\vec{X} = \vec{t} \wedge \vec{s})P(\vec{Y} = \vec{t} \vee \vec{s})$ holds for any $\vec{t}, \vec{s} \in \mathbb{R}^n$, then $\vec{X}$ is said to be smaller than or equal to $\vec{Y}$ with respect to the likelihood ratio order (for discrete random vectors) and it is denoted as $\vec{X} \leq_{lr} \vec{Y}$.*

It will also be of interest to consider some extensions of the usual stochastic order to other types of random elements. For stochastic processes, it is defined as the comparison of finite-dimensional distributions.

Definition 2.112 *[295] Let $(X_t, t \in T)$ and $(Y_t, t \in T)$ be two stochastic processes on the same index set. Then, if*

$$(X_{t_1}, \ldots, X_{t_n}) \leq_{st} (Y_{t_1}, \ldots, Y_{t_n}),$$

for any $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in T$, is said that $(X_t, t \in T)$ is smaller than or equal to $(Y_t, t \in T)$ in the usual stochastic order (for stochastic processes) and it is denoted by $(X_t, t \in T) \leq_{st} (Y_t, t \in T)$.

The reader is referred to Section 6.B.7 in [295] for some examples of stochastic processes ordered in this sense, including Markov and renewal processes.

The usual stochastic order for stochastic processes has an equivalent characterization as Condition 3 in Theorem 2.102. Similarly as for random vectors, the notation $(X_t, t \in T) \leq_{a.s.} (Y_t, t \in T)$ is used to denote $X_t \leq Y_t$ for all $t \in T$ with probability 1.

Theorem 2.113 *[295] Let $(X_t, t \in T)$ and $(Y_t, t \in T)$ be two stochastic processes on the same index set. Then, $(X_t, t \in T) \leq_{st} (Y_t, t \in T)$ if and only if there exist two stochastic processes $(\hat{X}_t, t \in T)$ and $(\hat{Y}_t, t \in T)$ such that $(\hat{X}_t, t \in T) =_{st} (X_t, t \in T)$, $(\hat{Y}_t, t \in T) =_{st} (Y_t, t \in T)$ and $(\hat{X}_t, t \in T) \leq_{a.s.} (\hat{Y}_t, t \in T)$.*

Moving to random sets, the closest notion of a location stochastic order that can be found in the literature is the paper by Montes et al. [242]. They provide several definitions of stochastic relations involving sets of distribution functions and p-boxes, relevant in the context of imprecise probabilities [236]. These relations can be straightforwardly adapted to random intervals. In the next definition, Definition 4 and Proposition 3 in [242] are partially adapted to introduce some of these relations for random intervals.

Definition 2.114 *[242] Let $I_X = [X_1, X_2]$ and $I_Y = [Y_1, Y_2]$ be two random intervals. Then, the following relations are defined:*

- *If $X_2 \leq_{st} Y_1$, then it is said that I_Y (FSD_1)-stochastically dominates I_X and it is denoted as $I_X \leq_{FSD_1} I_Y$,*
- *If $X_1 \leq_{st} Y_1$, then it is said that I_Y (FSD_2)-stochastically dominates I_X and it is denoted as $I_X \leq_{FSD_5} I_Y$,*
- *If $X_2 \leq_{st} Y_2$, then it is said that I_Y (FSD_5)-stochastically dominates I_X and it is denoted as $I_X \leq_{FSD_5} I_Y$.*

However, none of the relations are, at the same time, transitive, reflexive and antisymmetric with respect to the classes of equivalence related to having the same distribution (see Remark 1 in [242]), so they are not stochastic orders in the sense of Definition 2.100.

2.3.5.2 Variability stochastic orders

In some cases, it is not interesting to order the distributions in terms of location but to consider the variability. One of the classical variability stochastic comparisons is the dispersive order.

Definition 2.115 *[295] Let X and Y be two random variables with quantile functions F^{-1} and G^{-1}. Then, X is said to be smaller than or equal to Y in the dispersive order, denoted as $X \leq_{disp} Y$, if*

$$F^{-1}(p) - F^{-1}(q) \leq G^{-1}(p) - G^{-1}(q) \text{ for any } 0 < q \leq p < 1.$$

Note that $X \leq_{disp} Y$ holds if and only if the transformation $\phi : F^{-1} \circ G$, for which $X =_{st} \phi(Y)$, is increasing and fulfills

$$\phi(x) - \phi(y) \leq x - y, \text{ whenever } x \leq y,$$

with both $x, y \in \mathbb{R}$ (see Section 3.B in [295] for details). In addition, this is equivalent to requiring ϕ to be an increasing contraction (see Section 2.1.3).

The dispersive order implies the inequality of variances. Another common alternative is to use convexity to define stochastic comparisons based on variability. Between the most used, one can find convexity, componentwise convexity and symmetric convexity.

Definition 2.116 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors of dimension n. If $E[\varphi(\vec{X})] \leq E[\varphi(\vec{Y})]$ for any measurable convex (componentwise convex, symmetric convex) function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $E[\varphi(\vec{X})]$ and $E[\varphi(\vec{Y})]$ exist, then $\vec{X}$ is said to be smaller than or equal to $\vec{Y}$ in the convex (componentwise convex, symmetric convex) stochastic order and it is denoted as*

$$\vec{X} \leq_{cx} [\leq_{ccx}, \leq_{scx}] \vec{Y}.$$

Trivially, the componentwise convex order implies the convex order, which in turn implies both the symmetric convex and increasing convex orders. For random variables, all the orders introduced in Definition 2.116 are equivalent. In addition, if $E[X] = E[Y]$, then $X \leq_{disp} Y$ implies that $X \leq_{cx} Y$.

Since the functions $f(\vec{x}) = \pm x_i$ with $i \in [n]$ are convex, then $\vec{X} \leq_{cx} \vec{Y}$ implies that $E[\vec{X}] = E[\vec{Y}]$. Similarly, since $f(x) = x^2$ is convex, then $X \leq_{cx} Y$ implies that $\text{Var}(\vec{X}) \leq \text{Var}(\vec{Y})$. In addition, the convex order has similar characterizations to the ones for the increasing convex order in Theorem 2.109.

Theorem 2.117 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then, the following statements are equivalent.*

1. $\vec{X} \leq_{cx} \vec{Y}$,
2. *There exist two random vectors $\hat{\vec{X}}$ and $\hat{\vec{Y}}$ such that $\hat{\vec{X}} =_{st} \vec{X}$, $\hat{\vec{Y}} =_{st} \vec{Y}$ and $E\left[\hat{\vec{Y}} \,\middle|\, \hat{\vec{X}}\right] = \hat{\vec{X}}$.*

Moreover, if $\vec{X}$ and $\vec{Y}$ are random variables (denoted as X and Y), then the following statement is also equivalent.

3. $E[X] = E[Y]$ *and* $\int_x^\infty \bar{F}_X(t)dt \leq \int_x^\infty \bar{F}_Y(t)dt$, $\forall x \in \mathbb{R}$.

2.3.5.3 Dependence stochastic orders

For comparisons of random vectors in terms of dependence, different orders have been considered in the literature. One of the most used is the supermodular order. It is used to compare vectors whose distributions are members of the same Fréchet class.

Definition 2.118 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors that are members of the same Fréchet class. If $E[\varphi(\vec{X})] \leq E[\varphi(\vec{Y})]$ for any supermodular function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $E[\varphi(\vec{X})]$ and $E[\varphi(\vec{Y})]$ exist, then $\vec{X}$ is said to be smaller than or equal to $\vec{Y}$ in the supermodular order, and it is denoted as $\vec{X} \leq_{sm} \vec{Y}$.*

The supermodular order implies the inequality of the Pearson correlation coefficient. One can find a large number of results that describe conditions such that two random vectors are ordered with respect to the supermodular order. For example, results can be found for multivariate Gaussian distributions (Example 9.A.20 in [295]), exchangeable vectors with FGM copulas (Theorem 5 in [63]) or exchangeable multivariate Bernoulli random vectors (Theorem 2.11 [97]).

It is also a known fact that for bivariate random vectors, the supermodular order is equivalent to the PQD order [295], defined as follows.

Definition 2.119 *[295] Let $\vec{X}$ and $\vec{Y}$ be two random vectors that are in the same Fréchet class with distribution functions, respectively, F and G. Then, $\vec{X}$ is said to be smaller than or equal to $\vec{Y}$ in the positive quadrant dependent order, and it is denoted as $\vec{X} \leq_{PQD} \vec{Y}$ if $F(\vec{x}) \leq G(\vec{x})$ for any $\vec{x} \in \mathbb{R}^n$.*

When $n > 2$, the supermodular order implies the PQD order, but the other implication does not hold [295].

2.4 Statistics

Statistics refers to the field of Mathematics that deals with data analysis by using Probability Theory as a basis. It includes descriptive statistics, estimation of parameters, prediction of quantities or tests for hypothesis, among others methods.

The first use of the term is attributed to Girolamo Ghilini in 1633 and is derived from the Italian phrase *ragione di stato* [260]. Statistics, as a noun, was popularized by Gottfried Achenwall in 1749 [1]. Among other main contributors to Statistics, it is possible to find Pierre-Simon Laplace [317], Carl Friedrich Gauss [296], Karl Pearson [318], Ronald Fisher [202] and Jerzy Neyman [137].

In Statistics, data are associated with the observation of quantities of interest over a subset (the sample) of a bigger population of study. In this direction, Statistics can be separated into two big parts. The first one, Descriptive Statistics, focuses on summarizing and representing the given data without trying to conclude anything about the population. On the other hand, Statistical Inference tries to extrapolate the information from the sample to the population.

In particular, data are assumed to be observations of a random variable or vector that describes the behavior of the population. Therefore, Statistics can use the tools of Probability Theory in order to define and study the proposed methods.

For the purposes of this thesis, it is necessary to briefly introduce some statistical measures for samples, as well as the basics of the theory of estimation and hypothesis testing.

2.4.1 Statistical measures

In Descriptive Statistics, the values of a sample are usually summarized by computing the so-called statistical measures. In this section, the basic statistical measures regarding location and variability are given.

In the case of location, the main ones that are of interest for this thesis are the sample mean and the sample median. The first one is another name for the arithmetic mean (denoted by $\bar{x}$) of the sample, while the second one coincides with the central value of the data, when n is odd, and the average between the central values when n is even. Regarding variability measures, one of the most used is the sample variance [278], defined as

$$\hat{\sigma}^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{2n(n-1)}\sum_{i,j=1}^{n}(x_i-x_j)^2.$$

Another relevant variability measure is the sample Gini mean difference [303], which focuses on the absolute value of the differences in the sample,

$$G(\vec{x}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|.$$

Finally, another usual option is the range of the data [278], defined as

$$R(\vec{x}) = \max(\vec{x}) - \min(\vec{x}).$$

2.4.2 Estimation

In Statistics, it is common to have an expression for the distribution function of a random quantity X depending on some unknown parameters θ, which are associated with a particular statistical model. Denoting the set of possible values for the unknown parameters θ as Θ, an estimator of θ is a measurable function from the random sample to Θ that does not depend on the value of θ.

Definition 2.120 *[278] Consider $X_1(\theta),\ldots,X_n(\theta)$ identically distributed random variables with (marginal) distribution function F_θ depending on some unknown parameters $\theta \in \Theta$. An estimator is any measurable function $T : \mathbb{R}^n \to \Theta$ that does not depend on the value of the unknown parameters.*

The value $T(X_1(\theta),\ldots,X_n(\theta))$ takes for a particular observation of the sample is called an estimation of θ. Although many estimators can be defined for a set of parameters, there are desirable properties that estimators can fulfill. One of them is to require the estimator to be unbiased, i.e. the expected value of the estimator equals the real value of the parameter.

Definition 2.121 *[278] Consider $X_1(\theta),\ldots,X_n(\theta)$ identically distributed random variables with (marginal) distribution function F_θ depending on some unknown parameters $\theta \in \Theta$. An estimator $T : \mathbb{R}^n \to \Theta$ of θ is said to be unbiased if*

$$E[T(X_1(\theta),\ldots,X_n(\theta))] = \theta.$$

For instance, it is well-known that the arithmetic mean is an unbiased estimator of the expectation and the sample variance is an unbiased estimator of the variance. Another relevant property is efficiency, which focuses on the distance between the real value of the parameter and its estimation. In particular, it is defined in terms of the expectation of the square of the difference between the estimation and the real value of the parameters, i.e. $E\left[||T(X_1(\theta),\ldots,X_n(\theta)) - \theta||_2^2\right]$. This quantity is known as the Mean Squared Error (MSE).

Definition 2.122 *[278] Consider $X_1(\theta),\ldots,X_n(\theta)$ identically distributed random variables with (marginal) distribution function F_θ depending on some unknown parameters $\theta \in \Theta$ and let $T_1, T_2 : \mathbb{R}^n \to \Theta$ be two estimators of θ. Then, it is said that T_1 is more efficient than T_2 if*

$$E\left[||T_1(X_1(\theta),\ldots,X_n(\theta)) - \theta||_2^2\right] \leq E\left[||T_2(X_1(\theta),\ldots,X_n(\theta)) - \theta||_2^2\right], \forall \theta \in \Theta,$$

and there exists $\theta_0 \in \Theta$ such that

$$E\left[||T_1(X_1(\theta_0),\ldots,X_n(\theta_0)) - \theta_0||_2^2\right] < E\left[||T_2(X_1(\theta_0),\ldots,X_n(\theta_0)) - \theta_0||_2^2\right].$$

For example, the arithmetic mean is the most efficient estimator of the mean for Gaussian random variables [278]. The MSE, under some conditions, has a lower bound on the set of all estimators, known as the Frechet-Cramer-Rao bound [99, 130, 276]. The reader is referred to Section 8.3 in [278] for more properties such as sufficiency, completeness and consistence.

There is a direct connection between some aggregation functions and estimators of the centrality. The most relevant example is, again, the arithmetic mean as a mean estimator. Another prominent example is the use of OWA operators to estimate the expected value [72]. The reader is referred to some foundational papers [218, 287, 288, 286] and to some developments of such estimators in recent years [2, 121, 203].

Also related to estimation, the distribution function of a random variable can be approximated by, given a sample, computing the empirical distribution function. Roughly speaking, the empirical distribution function is the distribution function associated with choosing uniformly a value from the sample.

Definition 2.123 *[278] Let $X_1,\ldots,X_n$ be a sequence of random variables. The empirical distribution function is defined as:*

$$\hat{F}(x) = \begin{cases} 0 & \text{if } x < X_{(1)}, \\ \frac{k}{n} & \text{if } X_{(k)} \leq x < X_{(k+1)},\ k \in [n-1], \\ 1 & \text{if } X_{(n)} \leq x. \end{cases}$$

The empirical distribution function converges to the distribution function if the random variables have the same distribution and are independent. In addition, there are more involved results regarding the rate and the uniformity of the convergence, (see [257]).

2.4.3 Hypothesis testing

Another relevant part of Statistics is devoted to developing statistical tests in order to decide whenever a hypothesis can be assumed to be true or not. Given a random sample $(X_1,\ldots,X_n)$, a statistical test considers a null hypothesis and an alternative hypothesis. The random sample is assumed to fulfill the null hypothesis, and, under that assumption, the probability of having less or equally compatible samples is computed. This value is known as the p-value, and, if it is below a certain threshold known as significance level and usually considered as 0.05, the null hypothesis is rejected and the alternative accepted. The reader is referred to Chapter 3 in [181] or Chapter 9 in [278] for more information in this regard.

In the following, the statistical tests that will be relevant for the development of the thesis are briefly explained.

The one sample Kolmogorov-Smirnov test compares the empirical distribution function with a theoretical one by computing the supremum of the difference of both functions. The null hypothesis is that the sample follows a given theoretical distribution. Thus, small p-values will imply the rejection of that hypothesis [55]. Similarly, the two-sample Kolmogorov-Smirnov test is defined using two empirical distribution functions instead [55].

The one-sample Wilcoxon test, also known as the Mann-Whitney test, compares the number of values in the sample that are below and above a certain fixed value to determine whether that fixed value is central or not. The null hypothesis, in the case considered in this thesis, will be that the central location is equal to a particular value [321]. It is also possible to modify the test to compare the central value of two different populations.

The Cabilio-Masaro symmetry test is based on the difference between the sample mean and the sample median typified by the sample variance. If the difference is big enough, the null hypothesis of symmetry is rejected [77].

Another important problem is to discriminate whenever the usual stochastic order between two random variables holds, for which different hypothesis testing procedures have been developed [20, 232, 313]. In the case of the increasing convex order in the univariate case, there exist also several alternatives [18, 217, 333].

Chapter 3

Aggregation of random variables

Contents

In the study of the aggregation of random structures, the most simple case is to consider random variables. Aggregation of random variables is not only interesting because of its simplicity, it is also relevant due to the fact that, in most cases, random data take real numbers as values.

In this chapter, the basics of aggregations of random variables are disclosed. There are two main objectives. The first one is to study in detail the definition of aggregation of random variables, considering different alternatives. In addition, the scenarios that the considered definition can model are studied. The second objective is to provide a general framework that will serve as the theoretical basis for the studies and the results in the rest of the chapters.

In particular, the different possibilities for a definition of aggregations of random variables will be studied in Section 3.1. Using the adequate definition, some relevant families of aggregations of random variables are defined in Section 3.2. Finally, in Section 3.3 some additional properties of aggregations of random variables are introduced.

3.1 Definition of aggregation of random variables

As explained in Chapter 1, to the best knowledge of the author, there does not exist a notion of aggregation of random variables in the literature. Some particular works focused on usual aggregation functions with random inputs (see [117, 150, 193, 226, 227, 228, 230, 269]), but none of them establish a proper definition in probabilistic terms.

Therefore, the first step in the study of the aggregation of random variables is to construct the concept itself. This construction is the most important part of the thesis, since the concept is the basis of the rest of the results and particular cases. A non adequate notion of aggregation of random variables could lead to bad theoretical properties that would not allow the theory to be developed.

In this sense, the definition of aggregation of random variables has been refined in last years, mainly with respect to the boundary conditions. The first definition was given in [29], in which the boundary conditions were introduced in terms of the supremum and infimum of the set $L_I^n(\Omega)$ (which will be defined later). A detailed explanation of the boundary conditions was given later in [39]. Finally, the current form of the definition, see Definition 3.6, expresses the boundary conditions in a simple but equivalent manner and was introduced in [26].

The main notions about the definition of aggregations of random variables are addressed in this section. In particular, the construction process of such functions is described, explaining each of the necessary steps such as the construction of the domain and the image, the choice of the adequate stochastic order and the boundary conditions. Some direct implications of the definition are also provided.

3.1.1 A general definition

For defining aggregations of random variables, a reasonable technique is to try to extend Definition 2.5 to the probabilistic setting. In this regard, it is necessary to define the following three elements.

- A set of random vectors, related to an interval,
- A monotonicity condition,

- Boundary conditions.

The set of random vectors is easy to construct. Considering a probability space $(\Omega, \mathscr{F}, P)$ and a real interval I (bounded or not), consider the following set.

$$L_I^n(\Omega) = \left\{ \vec{X} : \Omega \to I^n \;\middle|\; \vec{X} \text{ is measurable} \right\}.$$

These sets are just the random vectors defined from the considered probability space to the Cartesian product of the interval I. If $n = 1$, $L_I^1(\Omega)$ will be denoted as $L_I(\Omega)$. In addition, (Ω) will be dropped whenever it does not lead to confusion, denoting them simply as L_I and L_I^n.

For monotonicity, it is necessary to have an order between random vectors. Of course, as introduced in Section 2.3.5, stochastic orders are defined to fulfill this purpose. In particular, it is natural to consider a location stochastic order, since they order random vectors with degenerate distribution just as the componentwise order orders usual vectors. In addition, it is necessary to have the stochastic order defined for any possible dimension of the random vectors. At this point, it will be considered a general location stochastic order $\leq_{so}$. Therefore, any candidate for being an aggregation of random variables A should satisfy $A(\vec{X}) \leq_{so} A(\vec{Y})$ if $\vec{X} \leq_{so} \vec{Y}$.

Finally, it is necessary to establish the boundary conditions. If the interval is bounded, i.e. $I = [a,b]$, then they can be defined by stating that $A(\vec{X}) =_{a.s.} a$ whenever $\vec{X} =_{a.s.} a\vec{1}$ and $A(\vec{X})_{a.s.} = b$ whenever $\vec{X} =_{a.s.} b\vec{1}$. If the interval is unbounded, the same reasoning as in Definition 2.6 can be applied. The set $L_{\mathbb{R}}$ can be extended with two new elements, $-\infty, \infty$ such that $-\infty \leq_{so} X \leq_{so} \infty$ for any $X \in L_{\mathbb{R}}$. Then, the considered boundary conditions will be the following.

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

Similarly to the case for degenerate vectors, the boundary conditions can be rewritten in a simpler form.

- For any $X \in L_I$, there exists $\vec{X} \in I$ such that $A(\vec{X}) \leq_{so} X$,

- For any $X \in L_I$, there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \geq_{so} X$.

Adding all these elements, aggregations of random variables can be defined as follows.

Definition 3.1 *[29] Let (Ω, Σ, P) be a probability space, $\leq_{so}$ a stochastic order and I a real non-empty interval. An aggregation function of random variables (with respect to $\leq_{so}$) is a function $A : L_I^n(\Omega) \to L_I(\Omega)$ which satisfies:*

1. *For any $\vec{X}, \vec{Y} \in L_I^n$ such that $\vec{X} \leq_{so} \vec{Y}$, $A(\vec{X}) \leq_{so} A(\vec{Y})$,*
2. *For any $X \in L_I$, there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \leq_{so} X$,*
3. *For any $X \in L_I$, there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \geq_{so} X$.*

3.1.2 On the choice of the stochastic order

In Definition 3.1, the aggregation of random variables with respect to a generic stochastic order is given. However, the properties of the notion are quite sensitive to the choice of such a stochastic order. If a strong stochastic order is chosen, then there are few pairs of $\vec{X}, \vec{Y} \in L_I^n$ such that $\vec{X} \leq_{so} \vec{Y}$, but is hard for $A(\vec{X}) \leq_{so} A(\vec{Y})$ to be fulfilled. On the contrary, if a weak stochastic order is chosen, the condition $A(\vec{X}) \leq_{so} A(\vec{Y})$ is easier to be verified, while the number of $\vec{X}, \vec{Y} \in L_I^n$ such that $\vec{X} \leq_{so} \vec{Y}$ increases.

Therefore, it is necessary to establish a criterion for choosing the stochastic order. Of course, it could depend on the particular case or field of application. However, a reasonable property that is generally desirable is to be able to use usual aggregation functions to define aggregations of random variables. That is, given an aggregation function $\hat{A} : I \to I^n$, it should be fulfilled that any $A : L_I^n \to L_I$ defined as $A(\vec{X}) = \hat{A} \circ \vec{X}$ is an aggregation of random variables with respect to the considered stochastic order.

In this direction, it is necessary to choose a stochastic order that is closed when an increasing function is applied. Specifically, it should hold that $\vec{X} \leq_{so} \vec{Y} \implies f(\vec{X}) \leq_{so} f(\vec{Y})$ for any measurable increasing function $f : \mathbb{R}^n \to \mathbb{R}$.

Sadly, there are many location stochastic orders that do not fulfill this property. This is the case, for instance, of the componentwise stochastic order and the Linear Stochastic Dominance.

Example 3.2 *[29] Consider two bivariate random vectors* $\vec{X} = (X_1, X_2)$ *and* $\vec{Y} = (Y_1, Y_2)$*, such that* $P(X_1 = 1, X_2 = 1) = P(X_1 = 0, X_2 = 0) = 0.5$ *and* $P(Y_1 = 0, Y_2 = 0) = 0.16$*,* $P(Y_1 = 0, Y_2 = 1) = P(Y_1 = 1, Y_2 = 0) = 0.24$ *and* $P(Y_1 = 1, Y_2 = 1) = 0.36$*. It is immediate that* $X_1 \leq_{st} Y_1$ *and* $X_2 \leq_{st} Y_2$*, so* $\vec{X} \leq_{cst} \vec{Y}$*. On the other hand, when aggregating the random vectors using the minimum function, it holds that* $P(\min(X_1, X_2) = 1) = 0.5$ *and* $P(\min(Y_1, Y_2) = 1) = 0.36$*. Thus,* $\min(\vec{X}) \not\leq_{cst} \min(\vec{Y})$.

Example 3.3 *[29] Consider two bivariate random vectors* $\vec{X} = (X_1, X_2)$ *and* $\vec{Y} = (Y_1, Y_2)$*, such that* $P(X_1 = 0, X_2 = 0) = P(X_1 = 0, X_2 = 0.5) = P(X_1 = 0.5, X_2 = 0) = P(X_1 = 0.5, X_2 = 0.5) = 0.25$ *and* $P(Y_1 = 0, Y_2 = 1) = P(Y_1 = 1, Y_2 = 0) = 0.5$*. Now, consider the positive linear combinations of the components* $\alpha X_1 + \beta X_2$ *and* $\alpha Y_1 + \beta Y_2$*, with* $\alpha, \beta \in \mathbb{R}^+$*. Notice that* $\alpha Y_1 + \beta Y_2$ *takes the values* α *and* β *with the same probability* 0.5 *and* $\alpha Y_1 + \beta Y_2$ *takes four equiprobable values, 0,* $\frac{\alpha}{2}$*,* $\frac{\beta}{2}$ *and* $\frac{\alpha+\beta}{2}$*. For any positive values of* α *and* β*,* $\alpha X_1 + \beta X_2 \leq_{st} \alpha Y_1 + \beta Y_2$ *always holds. Therefore,* $\vec{X} \leq_{LSD} \vec{Y}$*. On the other hand, when applying the minimum to both random vectors, one has* $P(\min(X_1, X_2) = 0) = 0.75$ *and* $P(\min(Y_1, Y_2) = 0) = 1$*. Thus,* $\min(\vec{X}) \not\leq_{LSD} \min(\vec{Y})$.

The intuition behind the reason why the latter stochastic orders are not suitable is immediate. The componentwise stochastic order does not keep into consideration the dependence between the random variables, while the Linear Stochastic Dominance only cares about the linear dependence. Therefore, the lack of consideration of the dependence structure of the random vectors does not allow one to have the order preserved when applying an increasing function.

Moving to stronger stochastic orders does not guarantee the preservation of the order either. For instance, some counterexamples can be found for the likelihood ratio order.

Example 3.4 *Let* $\vec{X} = (X_1, X_2)$ *and* $\vec{Y} = (Y_1, Y_2)$ *be two random vectors such that* $P(X_1 = 0, X_2 = 0) = P(X_1 = 0, X_2 = 1) = P(X_1 = 1, X_2 = 0) = P(X_1 = 1, X_2 = 1) =$

0.25 *and* $P(Y_1 = 0, Y_2 = 0) = 0$, $P(Y_1 = 1, Y_2 = 0) = 0.3$ *and* $P(Y_1 = 0, Y_2 = 1) = P(Y_1 = 1, Y_2 = 1) = 0.35$. *To prove that* $\vec{X} \leq_{lr} \vec{Y}$, *it is necessary* $P(\vec{X} = \vec{t})P(\vec{Y} = \vec{s}) \leq P(\vec{X} = \vec{t} \wedge \vec{s})P(\vec{Y} = \vec{t} \vee \vec{s})$ *to hold for any* $\vec{t}, \vec{s} \in \{0,1\}^2$. *If* $\vec{t} \leq \vec{s}$, *then the latter condition always holds. The rest of the cases can be found in Table 3.1, thus it is concluded that* $\vec{X} \leq_{lr} \vec{Y}$. *However, if the aggregation function defined as* $A(x_1, x_2) = 0.75x_1 + 0.25x_2$ *is considered, one has*

$$P(A(\vec{X}) = 0.75)P(A(\vec{Y}) = 0.25) = 0.25 \cdot 0.35 = 0.0875,$$
$$P(A(\vec{X}) = 0.25)P(A(\vec{Y}) = 0.75) = 0.25 \cdot 0.3 = 0.075,$$

and therefore, $A(\vec{X}) \not\leq_{lr} A(\vec{Y})$.

$\vec{t}$	$\vec{s}$	$P(\vec{X} = \vec{t})P(\vec{Y} = \vec{s})$	$\vec{t} \wedge \vec{s}$	$\vec{t} \vee \vec{s}$	$P(\vec{X} = \vec{t} \wedge \vec{s})P(\vec{Y} = \vec{t} \vee \vec{s})$
(1,0)	(0,0)	0	(0,0)	(1,0)	0.075
(1,0)	(0,1)	0.0875	(0,0)	(1,1)	0.0875
(0,1)	(0,0)	0	(0,0)	(0,1)	0.0875
(0,1)	(1,0)	0.075	(0,0)	(1,1)	0.0875
(1,1)	(0,0)	0	(0,0)	(1,1)	0.0875
(1,1)	(0,1)	0.0875	(0,1)	(1,1)	0.0875
(1,1)	(1,0)	0.075	(1,0)	(1,1)	0.0875

Table 3.1: Comparison of products of probabilities for the likelihood ratio order of Example 3.4.

Similar counterexamples, which are not going to be disclosed here, can be found for other location stochastic orders such as the hazard rate order, the weak hazard rate order or the orthant orders, whose definition can be found in [295].

Another possible option could be the Statistical Preference. However, it is not a stochastic order with respect to Definition 2.100, since it does not fulfill any of the required properties. It is easy to prove that the relation given in Definition 2.111 is preserved under the application of an increasing measurable function [29] and that the relation is complete for random variables. Considering other similar definitions (see [279]), it is possible to have the completeness also for random vectors, but by sacrificing the preservation with respect to increasing functions. Due to the

latter reasons, the Statistical Preference is not adequate to be used in Definition 3.1. However, it could be possible to work with Statistical Preference by considering the approach given in [266].

On the other hand, it is well-known that the usual stochastic order is closed under increasing functions (see Proposition 2.106). In addition, by applying Proposition 2.110, one has that for any measurable increasing function $f : \mathbb{R}^n \to \mathbb{R}$ and any random vectors $\vec{X}$ and $\vec{Y}$,

$$\vec{X} \leq_{ch} \vec{Y} \implies \vec{X} \leq_{st} \vec{Y} \implies f(\vec{X}) \leq_{st} f(\vec{Y}) \iff f(\vec{X}) \leq_{ch} f(\vec{Y}).$$

In this direction, the cumulative hazard order is preserved under the application of increasing functions, thus it is adequate for being $\leq_{so}$ in Definition 3.1. However, this stochastic order is stronger than the usual stochastic order, so its applicability is smaller. Moreover, it is only defined for positive random vectors [295], which is not always the case when dealing with real data. Another option is to impose some conditions about the copulas of the compared random vectors. In this direction, the same dependence structure usual stochastic order is defined.

Definition 3.5 *[39] Let $\vec{X}$ and $\vec{Y}$ be two random vectors with joint distribution functions F and G and marginal distributions $F_1, \ldots, F_n$ and $G_1, \ldots, G_n$. 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 $F_1(\mathbb{R}) \times \cdots \times F_n(\mathbb{R})$ such that*

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

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

Notice that if $\vec{X} \leq_{sd-st} \vec{Y}$, then $(F_1(X_1), \ldots, F_n(X_n)) =_{st} (G_1(Y_1), \ldots, G_n(Y_n))$ where $F_1, \ldots, F_n$ and $G_1, \ldots, G_n$ are the distribution functions of the components of, respectively, $\vec{X}$ and $\vec{Y}$ [39].

This stochastic order is stronger than the usual stochastic order, while it is equivalent for random variables. Therefore, it could be a choice for $\leq_{so}$ in Definition 3.1. However, similarly to the cumulative hazard order, since it is stronger than the usual stochastic order, its applicability is smaller.

The latter considerations are summarized in Table 3.2. It is concluded that the best option is to consider the usual stochastic order as the one that should be used to aggregate random variables. Then, the definition for this particular case will be the following.

Stochastic order	Monotonicity	Restrictions/Drawbacks
$\leq_{cst}$	No	No
$\leq_{LSD}$	No	No
$\leq_{lr}$	No	No
$\leq_{SP}$	Yes	It is not transitive
$\leq_{st}$	Yes	No
$\leq_{ch}$	Yes	Positive values
$\leq_{sd-st}$	Yes	Same dependence structure

Table 3.2: Monotonicity with respect to applying an increasing function and restrictions/drawbacks for the considered stochastic orders.

Definition 3.6 *[26, 29] Let (Ω,Σ,P) be a probability space and I a real non-empty interval. An aggregation function of random variables (with respect to $\leq_{st}$) is a function $A : L_I^n(\Omega) \to L_I(\Omega)$ which satisfies:*

1. *For any $\vec{X},\vec{Y} \in L_I^n$ such that $\vec{X} \leq_{st} \vec{Y}$, $A(\vec{X}) \leq_{st} A(\vec{Y})$,*
2. *For any $X \in L_I$, there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \leq_{st} X$,*
3. *For any $X \in L_I$, there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \geq_{st} X$.*

The latter definition will be considered the main one for the aggregation of random variables. That is, when referring to an aggregation of random variables, it will be assumed to be with respect to the usual stochastic order unless otherwise stated.

3.1.3 Some implications of the definition

Definition 3.6 is a non restrictive definition, as will be seen in Section 3.2, but imposes some immediate properties that any aggregation of random variables should

fulfill. The first one is that, if two random vectors have the same distribution, then the aggregated values also have the same distribution.

Lemma 3.7 *Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then, for any $\vec{X}, \vec{Y} \in L_I^n$ such that $\vec{X} =_{st} \vec{Y}$, $A(\vec{X}) =_{st} A(\vec{Y})$.*

Proof: Since $\vec{X} \leq_{st} \vec{Y}$, then $A(\vec{X}) \leq_{st} A(\vec{Y})$. Similarly, since $\vec{X} \geq_{st} \vec{Y}$, then $A(\vec{X}) \geq_{st} A(\vec{Y})$. It is concluded that $A(\vec{X}) =_{st} A(\vec{Y})$ ∎

Notice that this does not imply that $A(\vec{X}) =_{a.s.} A(\vec{Y})$ holds whenever $\vec{X} =_{a.s.} \vec{Y}$. Another important fact is that any aggregation of random variables leads to the definition of a function that can be seen as an aggregation of probability distributions. In this direction, the following set is introduced.

$$\mathscr{I}^n = \left\{ P : \mathbb{B}^n \to [0,1] \;\middle|\; \exists \vec{X} \in L_I^n \text{ such that } P_{\vec{X}} = P \right\},$$

where $\mathbb{B}^n$ denotes the Borel σ-algebra associated with $\mathbb{R}^n$. Their elements will be denoted as $P_{\vec{X}}$ with $\vec{X} \in L_I^n$. The notation $\mathscr{I}$ will be used for the case $\mathscr{I}^1$.

Proposition 3.8 *Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then, the function $A_d : \mathscr{I}^n \to \mathscr{I}$ given by $A_d(P_{\vec{X}}) = P_{A(\vec{X})}$ fulfills*

1. *For any $P_{\vec{X}}, P_{\vec{Y}} \in \mathscr{I}^n$ such that $P_{\vec{X}} \leq_{st} P_{\vec{Y}}$, $A_d(P_{\vec{X}}) \leq_{st} A_d(P_{\vec{Y}})$,*
2. *For any $P_X \in \mathscr{I}$, there exists $P_{\vec{X}} \in \mathscr{I}^n$ such that $A_d(P_{\vec{X}}) \leq_{st} P_X$,*
3. *For any $P_X \in \mathscr{I}$, there exists $P_{\vec{X}} \in \mathscr{I}^n$ such that $A_d(P_{\vec{X}}) \geq_{st} P_X$.*

Proof: Recall that $P_{\vec{X}}$ denotes the probability measures induced by the random vector $\vec{X}$ and, since Lemma 3.7 holds, A_d is well-defined. Notice that the usual stochastic order, as stated in Section 2.3.5, can be used both for random variables and probability measures over $\mathbb{B}^n$.

For monotonicity, since $P_{\vec{X}} \leq_{st} P_{\vec{Y}}$, then $\vec{X} \leq_{st} \vec{Y}$, which in turn implies that $A(\vec{X}) \leq_{st} A(\vec{Y})$. Therefore, $A_d(P_{\vec{X}}) = P_{A(\vec{X})} \leq_{st} P_{A(\vec{Y})} = A_d(P_{\vec{Y}})$.

For the boundary conditions, for any $X \in L_I$ there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \leq_{st} X$. Then, there exists $P_{\vec{X}} \in \mathscr{I}^n$ such that $A_d(P_{\vec{X}}) = P_{A(\vec{X})} \leq P_X$. For the upper bound, the proof is analogous. ∎

The natural question that arises is whenever any function $A_d(P_{\vec{X}}) = P_{A(\vec{X})}$ fulfilling the conditions established in Proposition 3.8 can be used as a way to define an aggregation of random variables. Although this case is more difficult, since every random vector has an associated probability measure but a probability measure can be associated with many random vectors, it is possible.

Proposition 3.9 *Let $A_d : \mathscr{I}^n \to \mathscr{I}$ be a function that fulfills conditions 1, 2 and 3 in Proposition 3.8. Consider a function $\phi : \mathscr{I} \to L_I$ such that $\phi(P_X) =_{st} X$ for any $X \in L_I$. Then, the function $A : L_I^n \to I$ given by $A(\vec{X}) = \phi(A_d(P_{\vec{X}}))$ for any $\vec{X} \in L_I^n$ is an aggregation of random variables.*

Proof: The function is well-defined since $P_{\vec{X}} \in \mathscr{I}^n$, $A_d : \mathscr{I}^n \to \mathscr{I}$ and $\phi : \mathscr{I} \to L_I$. For monotonicity, if $\vec{X} \leq_{st} \vec{Y}$, then $P_{\vec{X}} \leq_{st} P_{\vec{Y}}$, which in turn implies $A_d(P_{\vec{X}}) \leq_{st} A_d(P_{\vec{Y}})$ and $\phi(A_d(P_{\vec{X}})) \leq_{st} \phi(A_d(P_{\vec{Y}}))$. For the boundary conditions, for any $X \in L_I$ consider $P_{\vec{X}} \in \mathscr{I}^n$ such that $A_d(P_{\vec{X}}) \leq_{st} P_X$. Then, $\phi(A_d(P_{\vec{X}})) \leq_{st} \phi(P_X) =_{st} X$. For the upper bound, the proof is analogous. ∎

A way to construct the function ϕ will be provided in Proposition 3.20. Notice that this equivalence between the aggregation of random variables and the aggregation of probability measures over $\mathbb{R}^n$ is not possible just by using real functions applied to random vectors (see Section 7.6 in [292]).

When the interval is bounded, it is also possible to define a usual aggregation function by taking the expectation of the image of an aggregation of random variables over degenerate random vectors, as the following result states.

Proposition 3.10 *Let $I = [a,b]$ be a bounded real interval and let $A : L_I^n \to L_I$ be an aggregation of random variables. Then, the function $\hat{A} : I^n \to I$ defined as $\hat{A}(\vec{x}) = E[A(\vec{X})]$ with $\vec{X}$ being any random vector such that $\vec{X} =_{a.s.} \vec{x}$ is an aggregation function.*

Proof: For degenerate random vectors, it holds $\vec{X} =_{a.s.} \vec{Y} \iff \vec{X} =_{st} \vec{Y}$, therefore Lemma 3.7 ensures the same expectation for each possible choice of $\vec{X} =_{a.s.} \vec{x}$. In addition, it is clear that if $\vec{x} \in I^n$, then $\vec{X} \in L_I^n$ and if $X \in L_I$, $E[X] \in I$. Therefore, $\hat{A}$ is well-defined.

For monotonicity, if $\vec{x} \leq \vec{y}$, any random vectors $\vec{X}$ and $\vec{Y}$ such that $\vec{X} =_{a.s.} \vec{x}$ and $\vec{Y} =_{a.s.} \vec{y}$ fulfill $\vec{X} \leq_{a.s.} \vec{Y}$, thus $\vec{X} \leq_{st} \vec{Y}$. Then, $A(\vec{X}) \leq_{st} A(\vec{Y})$ and $\hat{A}(\vec{x}) = E[A(\vec{X})] \leq E[A(\vec{Y})] = \hat{A}(\vec{x})$.

For the boundary conditions, for any $x \in I$ one can consider $X =_{a.s.} x$ and then there exists $\vec{X}_1 \in L_I^n$ such that $A(\vec{X}_1) \leq_{st} X$. Moreover, one can consider $\vec{X}_2 =_{a.s.} a\vec{1}$ for which it is clear that $\vec{X}_2 \leq_{st} \vec{X}_1$. Then, $A(\vec{X}_2) \leq_{st} A(\vec{X}_1) \leq_{st} X$, thus $\hat{A}(a\vec{1}) = E[A(\vec{X}_2)] \leq E[A(\vec{X}_1)] \leq E[X] = x$. For the upper bound, the proof is analogous. ∎

A similar construction could be done using a fixed quantile of $A(\vec{X})$ instead of the mean. Unfortunately, for unbounded intervals, the latter result is not true, since, even if for any $X \in L_I$ there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \leq_{st} X$, there could not exist any choice of $\vec{X}$ with degenerate distribution. The construction of such aggregations of random variables is not intuitive, but a simple example can be provided.

Example 3.11 *Let $A : L_{\mathbb{R}}^n \to L_{\mathbb{R}}$ be the function defined as:*

$$A(X_1, X_2) = \begin{cases} X_1 & \text{if } S(X_1) \text{ has not a lower bound,} \\ \max(0, X_1) & \text{if } S(X_1) \text{ has a lower bound.} \end{cases}$$

The monotonicity is easy to verify, since x and $\max(x,0)$ are increasing, $x \leq \max(0,x)$ for any $x \in \mathbb{R}$ and if X_1 and X_2 are two random variables such that $S(X_1)$ has a lower bound and $S(X_2)$ has not a lower bound, it is clear that $X_1 \not\leq_{st} X_2$. For the lower boundary condition, for any $X \in L_{\mathbb{R}}$, if Y is an exponential random variable independent of X, it holds that $X - Y \leq_{st} X$. Moreover, since $S(X-Y)$ has not a lower bound, it holds that $A(X-Y, X-Y) = X - Y \leq_{st} X$. For the upper boundary condition, it is clear that $A(\max(0,X), X) = \max(0,X) \geq_{st} X$ with $(\max(0,X), X) \in L_{\mathbb{R}}^2$.

However, if one considers $X =_{a.s.} -1$, there is not any degenerate random vector (X_1, X_2) such that $A(X_1, X_2) \leq_{st} X$, since any degenerate random vector has a bounded support and then $A(X_1, X_2) = \max(0, X_1) \not\leq_{st} -1$.

The opposite problem, i.e. the construction of an aggregation of random variables by means of a usual aggregation function, will be studied in Section 3.2.1.

3.2 Families of aggregations of random variables

Although not evident at first glance, Definition 3.6 is a very flexible definition. In particular, it allows many types of functions to be considered as aggregations of random variables that go beyond the application of a fixed real function. This flexibility is caused mainly by the fact that the usual stochastic order, which focuses on the distributions but not the actual random vectors, is used to define the monotonicity and the boundary conditions. This has its own benefits and some drawbacks.

Firstly, the flexibility of the definition allows one to define 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 inputs and the output and to work with randomly behaved parameters [26]. As it will be seen in the subsequent sections, these particular cases have clear semantics and fields of application. They offer the possibility of considering other perspectives when aggregating data.

On the other hand, there is the possibility of having aggregations of random variables with non-intuitive definitions, without a clear semantic about the parameters or expressions and without applicability to data. For instance, recall the aggregation of random variables given in Example 3.11. This function fulfills all the axioms of being an aggregation of random variables, but its definition in terms of the boundaries of the supports of the random variables decreases its applicability to data. In general, one does not know about the existence or not of bounds of the support of the underlying random variable by just having some observations.

This section is devoted to exploring the limits of the definition of aggregation of random variables by defining families that are intuitive, with good properties and that focus on different probability concepts to define the aggregations. In addition, several results that increase the mathematical understanding of their structure are given.

3.2.1 Induced and degenerate aggregations

The most prominent, simple and intuitive type of aggregations of random variables one could think are usual aggregation functions applied to random vectors. That is,

aggregation functions with random inputs.

In this direction, the next result is one of the most relevant in order to have a coherent theory of aggregation of random variables. It will be called the Composition Theorem (of aggregation of random variables).

Theorem 3.12 *[29] Let $\hat{A} : I^n \to I$ be a measurable aggregation function. Then, the function $A : L_I^n \to L_I$ defined as $A(\vec{X}) = \hat{A} \circ \vec{X}$ is an aggregation of random variables.*

Proof: Notice that, for any $\vec{X} \in L_I^n$, $\vec{X} : \Omega \to I^n$ and, since $\hat{A} : I^n \to I$, then $\hat{A}(\vec{X}) : \Omega \to I$. The measurability of $\hat{A}(\vec{X})$ is a consequence of the measurability of both $\vec{X}$ and $\hat{A}$. Therefore, $\hat{A}(\vec{X}) \in L_I$ and A is well-defined.

For monotonicity, since the usual stochastic order is closed under increasing functions, see Proposition 2.106, then $\vec{X} \leq_{st} \vec{Y}$ implies $\hat{A}(\vec{X}) \leq_{st} \hat{A}(\vec{Y})$.

For the boundary conditions, suppose that I has a lower bound a. If $\vec{X} =_{st} (a, \ldots, a)$, by the boundary conditions of $\hat{A}$, one has $\hat{A}(\vec{X}) =_{st} a$.

If I has not a lower bound, consider $X \in L_I$. For each integer $k \in I \cap \mathbb{Z}$, consider $\vec{x}_k \in I^n$ such that $\hat{A}(\vec{x}_k) < k$, which exists by the boundary conditions of $\hat{A}$. Then, define the function $h : I \to I^n$ as $h(x) = \vec{x}_k$ if $x \in [k, k+1) \cap I$. Notice that, since $[k, k+1) \cap I$ is a measurable set for any $k \in \mathbb{Z}$, h is a measurable function. In addition, since its image is I^n, one has $h(X) \in L_I^n$. By construction, $A(h(X)) \leq_{a.s.} X$ and it is concluded that for any $X \in L_I$ there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \leq_{st} X$.

For the upper bound, the proof is analogous. It is concluded that A is an aggregation of random variables. ∎

Notice that the aggregation function is required to be measurable, a property that may not be satisfied. More in detail, any increasing function from $\mathbb{R}^n$ to $\mathbb{R}$ is Lebesgue measurable (see Theorem 4.4 in [152]), which is not enough since (Borel) measurability is needed. However, the examples of increasing functions that are not measurable are pathological and uncommon, mainly defined artificially by means of non-measurable sets, (see again [152]). In fact, all the aggregation functions introduced in Section 2.2.3 are measurable.

Throughout the rest of the thesis, this type of aggregations of random variables will be called induced.

Definition 3.13 *[29] Let $A : L_I^n \to L_I$ be an aggregation of random variables. If there exists an aggregation function $\hat{A} : I^n \to I$ such that $A(\vec{X}) = \hat{A} \circ \vec{X}$ for any $\vec{X} \in L_I^n$, it is said that A is induced (by $\hat{A}$).*

There are as many induced aggregations of random variables as there are measurable aggregation functions. For instance, those induced by the arithmetic mean or OWA operators are widely used as estimators in Statistics [72, 278].

Another simple family of aggregations of random variables are the degenerate ones, for which the image can always be associated with a non-random real number.

Definition 3.14 *[26, 31] Let $A : L_I^n \to L_I$ be an aggregation of random variables. If $A(\vec{X})$ has degenerate distribution for any $\vec{X} \in L_I^n$, A is said to be degenerate.*

An easy way to define a degenerate aggregation of random variables is to use the expectation or the quantiles to the different random variables. For instance, it can be proved that

$$A(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} E[X_i],$$

is a degenerate aggregation of random variables if the interval I is bounded. Otherwise, it is not. Moreover, there does not exist any degenerate aggregation of random variables over unbounded intervals.

Proposition 3.15 *Let I be an unbounded interval. Then, there does not exist any degenerate aggregation of random variables $A : L_I^n \to L_I$.*

Proof: Consider X to be a random variable with support I. Then, since for any $\vec{X} \in L_I^n$ the random variable $A(\vec{X})$ has degenerate distribution, it holds that neither $A(\vec{X}) \leq_{st} X$ nor $A(\vec{X}) \geq_{st} X$ can hold. Therefore, the boundary conditions are not fulfilled. ∎

These aggregations appear in prediction problems, in which in the last step of the process a single value (not a distribution) should be used as the prediction. For instance, a conditional expected value is often used as a point prediction in time series models such as, for instance, the ARMA model, see Chapter 3 in [156].

3.2.2 Conditionally determined aggregations

When using an aggregation of random variables in applied prediction or estimation problems, a desirable property is to be able to compute the value of the aggregation given an observation of the input random vector. Any aggregation of random variables that fulfills this property will be called conditionally determined. Working with conditional distributions, this is equivalent to requiring the conditional distribution of $A(\vec{X})$ given any value of $\vec{X}$ to be degenerate.

Definition 3.16 *[26] 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 $\vec{X} \in L_I^n$ and $\vec{x} \in I^n$ for which $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ is well defined, $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ 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 vector and its value. Notice that, unlike induced aggregations, the value could change depending on the input random vector. In the following examples, two conditionally determined aggregations are provided.

Example 3.17

- *[26] Given a fixed value $z \in I$, let $A : L_I^n \to L_I$ be the aggregation of random variables given by $A(\vec{X}) = \max(\vec{X})$ if $\vec{X} \leq_{st} z\vec{1}$, $A(\vec{X}) = \min(\vec{X})$ if $\vec{X} \geq_{st} z\vec{1}$ and $A(\vec{X}) = z\vec{1}$ otherwise. A, which can be seen as a generalization of nullnorms, is conditionally determined,*

- *Let $A : L_I^n \to L_I$ be the aggregation of random variables given by $A(X_1, X_2) = \frac{1}{2}E[X_1] + \frac{1}{2}X_2$. A, which is the average between the expectation of the first variable and the value of the second, is conditionally determined.*

Conditional determination appears naturally in some areas of Statistics. For mean estimation, the function applied to the random sample varies depending on the distribution. For instance, the most efficient estimator for Gaussian distributions is the arithmetic mean and, for uniform distributions, is the average between the maximum and the minimum [278]. In this direction, conditionally determined aggregations allow one to introduce prior information of the distribution of the aggregated data into the aggregation process. In applications, such information can

be estimated using prior data. The reader is referred to Sections 4.4.3 and 7.4 for the definition and application in a prediction problem of a conditionally determined aggregation based on distribution functions. In the following result, two alternative characterizations of this type of aggregations of random variables are given.

Theorem 3.18 *[26] Let $A : L_I^n \to L_I$ be an aggregation of random variables. The following statements are equivalent.*

(1) A is conditionally determined,

(2) For any $\vec{X} \in L_I^n$, there exists $Y \in L_I$ such that $Y =_{a.s.} A(\vec{X})$ and Y is measurable with respect to $\sigma(\vec{X})$,

(3) There exists a measurable family of functions $\left(G_{\vec{X}}, \vec{X} \in L_I^n\right)$ fulfilling $A(\vec{X}) =_{a.s.} G_{\vec{X}} \circ \vec{X}$ for any $\vec{X} \in L_I^n$.

Proof: Suppose that A fulfills (1). For any $\vec{X} \in L_I^n$, let $C_{\vec{X}} \subseteq \mathbb{R}^n$ be the set for which $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ is well-defined for any $\vec{x} \in C_{\vec{X}}$ and $P(\vec{X} \in C_{\vec{X}}) = 1$, which exists by Theorem 2.63. Define $G_{\vec{X}} : I^n \to I$ as the function such that $G_{\vec{X}}(\vec{x}) = \lambda$ with λ the value that $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ takes with probability 1 if $\vec{x} \in C_{\vec{X}}$ and $G_{\vec{X}}(\vec{x}) = 0$ otherwise. Since $P(\vec{X} \in C_{\vec{X}}) = 1$, it is concluded that $A(\vec{X}) =_{a.s.} G_{\vec{X}} \circ \vec{X}$ for any $\vec{X} \in L_I^n$. For measurability, notice that $G_{\vec{X}}(\vec{x}) = E[A(\vec{X}) \mid \vec{X} = \vec{x}]$ if $\vec{x} \in C_{\vec{X}}$ and $G_{\vec{X}}(\vec{x}) = 0$ otherwise and that the conditional expectation is measurable, see Definition 2.64. Then, (3) holds.

Suppose that (3) holds. For any $B \in \mathbb{B}$, $(G_{\vec{X}}(\vec{X}))^{-1}(B) = \vec{X}^{-1}\left(G_{\vec{X}}^{-1}(B)\right) \in \sigma(\vec{X})$. Then, (2) holds.

Suppose that (2) holds. Then, given $\vec{X} = \vec{x}$, one has that $\vec{X}^{-1}(\vec{x})$ is a measurable set that does not contain any other measurable set (in $\sigma(\vec{X})$). Then, Y should take a unique value on $\vec{X}^{-1}(\vec{x})$. That is, $[Y \mid \vec{X} = \vec{x}]$ has degenerate distribution if well-defined. Since $A(\vec{X}) =_{a.s.} Y$, the same holds for $A(\vec{X})$. Then, A is conditionally determined and (1) is fulfilled. ∎

The third point in the last result gives a comprehensible characterization of conditional determination. Firstly, the random vector $\vec{X}$ is identified and then a

particular function, $G_{\vec{X}}$, is selected and applied to $\vec{X}$ to obtain the output random variable. Of course, not all choices of $\left(G_{\vec{X}}, \vec{X} \in L_I^n\right)$ are suitable for defining an aggregation of random variables. Measurability, monotonicity and the boundary conditions should be guaranteed.

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

3.2.3 Randomly induced aggregations

Conditional determination imposes that the values that the input random vector takes allow one to compute the value of the output random variable. However, there are cases where this is not true. It is possible to have randomness in the aggregation process itself, leading to having a non determined output even with determined inputs. In this direction, the concept of random aggregation of random variables is introduced as a negation of conditional determination.

Definition 3.19 *[26, 31] Let $A : L_I^n \to L_I$ be an aggregation of random variables. It is said that A is random if it is not conditionally determined.*

That is, A is random if there exist $\vec{X} \in L_I^n$ and $\vec{x} \in I^n$ such that the conditional distribution of $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ is well-defined and non-degenerate [31]. One particular example of interest is the construction of aggregations of random variables by aggregations of distributions (see Proposition 3.9). In the following, a particular case that considers an internal continuous aggregation function over $[0,1]$ is provided.

Proposition 3.20 *[33] Let I be a real interval, U a standard uniform random variable and $\hat{A} : [0,1]^n \to [0,1]$ an internal continuous aggregation function. Then, if for each random vector $\vec{X} = (X_1, \ldots, X_n)$ with marginal distribution functions $F_1, \ldots, F_n$ the function $A : L_I^n \to L_I$ is defined as*

$$A(\vec{X}) = \left(\hat{A}(F_1, \ldots, F_n)\right)^{-1}(U),$$

then A is an aggregation of random variables.

Proof: Since $\hat{A}$ is an aggregation function, $\hat{A}(F_1,\dots,F_n)$ is increasing. Moreover, since $\hat{A}$ is continuous, $\hat{A}(F_1,\dots,F_n)$ is right-continuous, the limit in $-\infty$ is 0 and the limit in ∞ is 1. Therefore, $\hat{A}(F_1,\dots,F_n)$ is a distribution function. In particular, it is the distribution function of $A(\vec{X})$.

For monotonicity, let $\vec{X}$ and $\vec{Y}$ be two random vectors with marginal distribution functions $F_1,\dots,F_n$ and $G_1,\dots,G_n$ such that $\vec{X} \leq_{st} \vec{Y}$. Then, the marginals are also ordered with respect to the usual stochastic order, so $F_i(x) \geq G_i(x)$ for each $x \in \mathbb{R}$ and $i \in [n]$. Then, since $\hat{A}$ is increasing, one has $\hat{A}(F_1(x),\dots,F_n(x)) \geq \hat{A}(G_1(x),\dots,G_n(x))$ for all $x \in \mathbb{R}$. Therefore, it is concluded that $A(\vec{X}) \leq_{st} A(\vec{Y})$.

For the boundary conditions, if the interval I has a lower bound a, then it is clear that $A(a,\dots,a) = a$ and the boundary conditions are fulfilled. If not, let $\vec{X} \in L_I$ with distribution function F. Then, consider the random vector $(X,\dots,X)$. Then, it is clear that $A(\vec{X})$ has distribution function $\hat{A}(F,\dots,F) = F$ (since $\hat{A}$ is idempotent, see Proposition 2.10). Therefore, $A(\vec{X}) =_{st} X$ and both boundary conditions are met. ■

This type of aggregations of random variables will be called distribution-based (on $\hat{A}$). Notice that, if $\hat{A}$ is a weighted arithmetic mean, then the latter aggregation function is equivalent to a mixture. Continuity of $\hat{A}$ is a necessary condition, since $\hat{A}(F_1,\dots,F_n)$ needs to be a distribution function. If it is not continuous, then $\hat{A}(F_1,\dots,F_n)$ could not be right-continuous or its limit $\lim_{x\to\infty}\hat{A}(F_1(x),\dots,F_n(x))$ could not be 1. However, internality is not necessary. For instance, the product can be used as the aggregation $\hat{A}$. Unfortunately, a general proof for the case of non-internality is not known, since the boundary conditions for the unbounded case are hard to prove.

Leaving aside distribution-based aggregations of random variables, random aggregations of random variables is a huge family that includes many useless and counterintuitive examples. In the following, a useful and understandable particular case is studied, the aggregations with random parameters.

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, Chapter 5 in [170], Section 5.2 in [49] or [21, 22]). If the training data are considered as realizations of random variables, then the fitted parameters also have a random behavior. In addition, there are sce-

narios in which the theoretical parameters are considered to be random (see [14]).

In the construction of aggregations of random variables with random parameters, one may think of modeling the random parameters by considering a random vector. However, this choice is not adequate in terms of monotonicity, as illustrated in the following example.

Example 3.21 *[26] 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 arithmetic 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 latter 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 this 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 3.22 *[26] Let I be a real interval, $\hat{A} : I^n \times \mathbb{R}^d \to I$ a measurable function and $(\vec{\lambda}_{\vec{X}}, \vec{X} \in L_I^n)$ a family of random vectors such that:*

(1) For any $\vec{z} \in \mathbb{R}^d$, the function $\hat{B}_{\vec{z}} : I^n \to I$ defined as $\hat{B}_{\vec{z}}(x_1, \ldots, x_n) = \hat{A}(x_1, \ldots, x_n, z_1, \ldots, z_d)$ is an aggregation function,

(2) If I does not have a lower (upper) bound, for any $x \in I$ there exists $\vec{x} \in I^n$ such that $\hat{A}(\vec{x}, \vec{z}) < (>) x$ for any $\vec{z} \in \mathbb{R}^d$,

(3) $\vec{\lambda}_{\vec{X}}$ has the same distribution for any $\vec{X} \in L_I^n$,

(4) For any $\vec{X}, \vec{Y} \in L_I^n$ such that $\vec{X} \leq_{st} \vec{Y}$, $[\vec{X} \mid \vec{\lambda}_{\vec{X}} = \vec{z}] \leq_{st} [\vec{Y} \mid \vec{\lambda}_{\vec{Y}} = \vec{z}]$ for any $\vec{z} \in S(\vec{\lambda}_{\vec{X}})$.

Then, the function $A : L_I^n \to L_I$ defined as $A(\vec{X}) = \hat{A}(\vec{X}, \vec{\lambda}_{\vec{X}})$ 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 monotonicity, consider $\vec{X}, \vec{Y} \in L_I^n$ such that $\vec{X} \leq_{st} \vec{Y}$. Notice that, since $\vec{\lambda}_{\vec{X}}$ and $\vec{\lambda}_{\vec{Y}}$ have the same distribution, then $S(\vec{\lambda}_{\vec{X}}) = S(\vec{\lambda}_{\vec{Y}})$. Using that $\hat{B}_{\vec{z}}$ is increasing for any $\vec{z} \in \mathbb{R}^d$, $\hat{B}_{\vec{z}}([\vec{X} \mid \vec{\lambda}_{\vec{X}} = \vec{z}]) \leq_{st} \hat{B}_{\vec{z}}([\vec{Y} \mid \vec{\lambda}_{\vec{Y}} = \vec{z}])$ for any $\vec{z} \in S(\lambda_{\vec{X}})$. Finally, since $\lambda_{\vec{X}} =_{st} \lambda_{\vec{Y}}$, applying Proposition 2.104 and 3 in Theorem 2.102, one has that $A(\vec{X}) = \hat{A}(\vec{X}, \vec{\lambda}_{\vec{X}}) = \hat{B}_{\vec{\lambda}_{\vec{X}}}(\vec{X}) \leq_{st} \hat{B}_{\vec{\lambda}_{\vec{Y}}}(\vec{Y}) = \hat{A}(\vec{Y}, \vec{\lambda}_{\vec{Y}}) = A(\vec{Y})$.

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

If I has not a lower bound, consider $X \in L_I$. For each $k \in I \cap \mathbb{Z}$, consider $\vec{x}_k \in I^n$ such that $\hat{A}(\vec{x}, \vec{z}) < x$ for any $\vec{z} \in \mathbb{R}^d$, which exists by hypothesis. Then, define the function $h : I \to I^n$ as $h(x) = \vec{x}_k$ 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, $A(h(X)) \leq_{a.s.} X$ and it is concluded that for any $X \in L_I$ there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \leq_{st} X$. For the upper bound, the proof is the same. ∎

An aggregation of random variables constructed as in Theorem 3.22 will be said to be randomly induced (by $\hat{A}$ and $(\vec{\lambda}_{\vec{X}}, \vec{X} \in L_I^n)$). This result can be seen as a Composition Theorem for randomly induced aggregations of random variables, since it generalizes Theorem 3.12.

The conditions imposed over $(\vec{\lambda}_{\vec{X}}, \vec{X} \in L_I^n)$ might seem very strong, but a simple case in which they hold is when $\vec{X}$ and $\vec{\lambda}_{\vec{X}}$ are independent for any $\vec{X} \in L_I^n$. Notice that these types of structures can be defined in a probability space that fulfills the saturation property introduced in Definition 2.44. More involved cases maybe can be found by fixing a vector copula (see [125]) between $\vec{X}$ and $\vec{\lambda}_{\vec{X}}$. For the second condition of the theorem, it is enough to consider, for instance, the aggregation $\hat{B}_{\vec{z}}$ to be internal for any $\vec{z} \in \mathbb{R}^d$.

Using Theorem 3.22 it is possible to overcome the problems in Example 3.21 to define aggregations of random variables with random parameters.

Example 3.23 *[31] Let $(U_{\vec{X}}, \vec{X} \in L_{\mathbb{R}}^n)$ be a family of standard uniform random variables such that $U_{\vec{X}}$ is independent of $\vec{X}$ for any $\vec{X} \in L_{\mathbb{R}}^n$. Consider the function $A : L_{\mathbb{R}}^n \to L_{\mathbb{R}}$ that takes a random vector $\vec{X}$ and returns the random variable*

$$A(\vec{X}) = \Big(\max(\vec{X}) - \min(\vec{X})\Big) U_{\vec{X}} + \min(\vec{X}).$$

Then, using Theorem 3.22, A is an aggregation of random variables. This aggregation can be seen as a generalization of the estimation procedure for a uniform random variable. Given a vector of independent and identically distributed random variables $X_1, \ldots, X_n$, the maximum likelihood estimators of the parameters of a uniform distribution are the minimum and the maximum of the sample [72, 278]. If the uniform distribution is constructed using the estimations, it coincides with $A(\vec{X})$. Notice that, in this case, general random vectors and not only independent and identically distributed random variables can be considered.

3.2.4 Aggregations with the same distribution

With a closer look at Definition 3.6, the monotonicity focuses on the distribution of $\vec{X}$ and $A(\vec{X})$, but not on the dependence between them. In particular, monotonicity imposes that $\vec{X} =_{st} \vec{Y}$ implies $A(\vec{X}) =_{st} A(\vec{Y})$, as already proved in Lemma 3.7, but it does not impose neither that $\vec{X} =_{st} \vec{Y}$ implies $(\vec{X}, A(\vec{X})) =_{st} (\vec{Y}, A(\vec{Y}))$ nor that $\vec{X} =_{a.s.} \vec{Y}$ implies $A(\vec{X}) =_{a.s.} A(\vec{Y})$ (see Section 3.3.3 for some extensions of monotonicity in this direction). This allows some particular types of aggregation of random variables to be defined.

Example 3.24 *[26] Let $A : L_I^n \to L_I$ be an induced aggregation. Consider U a uniform random variable. For any $\vec{X} \in L_I^n$, define $B(\vec{X}) = F^{-1}_{A(\vec{X})}(U)$, where $F^{-1}_{A(\vec{X})}$ is the quantile function of $A(\vec{X})$. Trivially, $B(\vec{X}) =_{st} A(\vec{X})$. Therefore, $B : L_I^n \to L_I$ is an aggregation of random variables. However, B is not induced. Moreover, B is random, because the value of the output random variable depends on U but not necessarily on the value of $\vec{X}$.*

In the latter example, B is 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. The next definition formalizes the idea.

Definition 3.25 *[26] Let $A,B : L_I^n \to L_I$ be two aggregations of random variables. If $A(\vec{X}) =_{st} B(\vec{X})$ for any $\vec{X} \in L_I^n$, it is said that A and B have the same distribution.*

Unfortunately, constructing an aggregation of random variables with the same distribution as another one is not interesting, in general, since its applicability is quite limited. However, a particular type, the aggregations with the same distribution as a conditionally determined one, can be characterized in terms of measure preserving functions and appear in some contexts regarding time series. In the following result, the extension of the family of conditionally determined aggregations of random variables is studied by considering the aggregations that have the same distribution.

Theorem 3.26 *[26] Let $(\Omega, \mathscr{F}, P)$ be a hyperfinite probability space and let $A : L_I^n(\Omega) \to L_I(\Omega)$ 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 $\left(\phi_{\vec{X}}, \vec{X} \in L_I^n\right)$ and random vectors $\left(\vec{Z}_{\vec{X}}, \vec{X} \in L_I^n\right)$ such that $A(\vec{X}) \circ \phi_{\vec{X}} =_{a.s.} \vec{Z}_{\vec{X}}$ and $\vec{Z}_{\vec{X}}$ is measurable with respect to $\sigma(\vec{X})$ for any $\vec{X} \in L_I^n$,

(3) For any $\vec{X} \in L_I^n$, if $S_{\vec{X}}$ denotes the set of probability of mass points of $\vec{X}$ and $S_{A(\vec{X})}$ denotes the probability of mass points of $A(\vec{X})$, there exists $L : S_{\vec{X}} \to S_{A(\vec{X})}$ such that:

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

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

Proof: Suppose that (1) holds. Then, there exists a conditionally determined aggregation of random variables $B : L_I^n \to L_I$ such that $B(\vec{X}) =_{st} A(\vec{X})$ for any $\vec{X} \in L_I^n$. Then, (2) holds by using Theorem 2.43 and $(1) \Rightarrow (2)$ in Theorem 3.18.

Suppose that (2) holds. For any $\vec{X} \in L_I^n$, since $\phi_{\vec{X}}$ is a measure preserving transformation, $A(\vec{X}) =_{st} A(\vec{X}) \circ \phi_{\vec{X}}$. Applying $(2) \Rightarrow (1)$ in Theorem 3.18, $A(\vec{X}) \circ \phi_{\vec{X}}$ is conditionally determined and (1) holds.

Suppose that (1) holds. Applying (1) $\Rightarrow$ (3) in Theorem 3.18, for any $\vec{X} \in L_I^n$, there exists a measurable function $G_{\vec{X}}$ such that $G_{\vec{X}}(\vec{X}) =_{st} A(\vec{X})$. Define $L : S_{\vec{X}} \to S_{A(\vec{X})}$ as $L(\vec{x}) = G_{\vec{X}}(\vec{x})$. Then,

$$P(A(\vec{X}) = x) = P\left(\vec{X} \in G_{\vec{X}}^{-1}(x)\right) \geq P\left(\vec{X} \in L^{-1}(x)\right) = \sum_{\vec{x} \in L^{-1}(x)} P(\vec{X} = \vec{x}),$$

for any $x \in S_{A(\vec{X})}$. It is concluded that (3) holds.

Suppose that (3) holds. Decompose the distribution function of $\vec{X}$ as $F_{\vec{X}} = \lambda F_{\vec{X} \mid \vec{X} \in S_{\vec{X}}} + (1-\lambda) F_{\vec{X} \mid \vec{X} \notin S_{\vec{X}}}$, where $\lambda = P(\vec{X} \in S_{\vec{X}})$. Notice that $F_{\vec{X} \mid \vec{X} \notin S_{\vec{X}}}$ is continuous (not necessarily absolutely continuous), so $F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}$ is continuous. Consider the following function,

$$F_0 = \left(F_{A(\vec{X})} - \lambda F_{L(\vec{X}) \mid \vec{X} \in S_{\vec{X}}}\right) \frac{1}{1-\lambda}.$$

Trivially, F_0 is right-continuous, $\lim_{x \to -\infty} F_0(x) = 0$ and $\lim_{x \to \infty} F_0(x) = 1$. Moreover, since $\sum_{\vec{x} \in L^{-1}(x)} P(\vec{X} = \vec{x}) \leq P(A(\vec{X}) = x)$, F_0 is increasing, thus is a distribution function.

Consider the function $G_{\vec{X}} : \mathbb{R}^n \to \mathbb{R}$ defined as:

$$G_{\vec{X}}(\vec{x}) = \begin{cases} L(\vec{x}) & \text{if } \vec{x} \in S_{\vec{X}}, \\ F_0^{-1}\left(F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}(x_1)\right) & \text{otherwise.} \end{cases}$$

The function $G_{\vec{X}}$ is measurable, since $S_{\vec{X}}$ is countable and $F_0^{-1} \circ F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}$ is measurable. Moreover, if one computes the distribution function of $G_{\vec{X}}(\vec{X})$:

$$\begin{aligned} F_{G_{\vec{X}}(\vec{X})} &= \lambda F_{G_{\vec{X}}(\vec{X}) \mid \vec{X} \in S_{\vec{X}}} + (1-\lambda) F_{G_{\vec{X}}(\vec{X}) \mid \vec{X} \notin S_{\vec{X}}} = \\ &= \lambda F_{L(\vec{X}) \mid \vec{X} \in S_{\vec{X}}} + (1-\lambda) F_{F_0^{-1}\left(F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}(X_1 \mid \vec{X} \notin S_{\vec{X}})\right) \,\middle|\, \vec{X} \notin S_{\vec{X}}} = \\ &= \lambda F_{L(\vec{X}) \mid \vec{X} \in S_{\vec{X}}} + (1-\lambda) F_0 = \\ &= \lambda F_{L(\vec{X}) \mid \vec{X} \in S_{\vec{X}}} + \left(F_{A(\vec{X})} - \lambda F_{L(\vec{X}) \mid \vec{X} \in S_{\vec{X}}}\right) \frac{1-\lambda}{1-\lambda} = F_{A(\vec{X})}. \end{aligned}$$

where it has been used that, since $F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}$ is continuous, $F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}(X_1 \mid \vec{X} \notin S_{\vec{X}})$ has standard uniform distribution and, therefore, $F_0^{-1}(F_{X_1 \mid \vec{X} \notin S_{\vec{X}}}(X_1 \mid \vec{X} \notin S_{\vec{X}}))$ has

distribution function F_0. It is concluded that there exists a family of measurable functions $(G_{\vec{X}}, \vec{X} \in L_I^n)$ such that, for any $\vec{X} \in L_I^n$, $G_{\vec{X}}(\vec{X}) =_{st} A(\vec{X})$ and, applying $(3) \Rightarrow (1)$ in Theorem 3.18, that (1) holds. ∎

The change in 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 by Ergodic Theory [268]. Any real-valued stationary process over the integers is associated with a measure preserving transformation (see 1.3 (g) in [101] or 1.2.B in [268]). In particular, if $(X_n, n \in \mathbb{N})$ is a stationary stochastic process, there exists a measure-preserving transformation $\phi : \Omega \to \Omega$ such that $X_{n+1} = X_n \circ \phi$ for any $n \in \mathbb{N}$. Therefore, an induced aggregation of random variables applied to $X_k, \ldots, X_{k+n-1}$ and $X_{k+1}, \ldots, X_{k+n}$ produces two different outputs with the same distribution that are linked by the measure preserving transformation ϕ.

Informally, the last characterization in Theorem 3.26 says that if it is possible to find functions that fit the probability mass points of the inputs in the probability mass points of the outputs, then the aggregation of random variables has the same distribution as a conditionally determined one. The most simple example of a scenario in which this does not happen is when a degenerate random vector has associated an aggregated random variable that is continuous.

3.2.5 Relationship between the families

Given the definitions of the different families, this section is devoted to studying the relationship between them. Leaving aside the trivial fact that random and conditionally determined aggregations are two disjoint subsets that cover all possible aggregations of random variables, all the relations between the families are stated in the next result.

Proposition 3.27 *[26, 29] Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then,*

(1) If I consists in more than one real number, then A cannot be induced and degenerate,

(2) 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,

(3) 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,

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

(5) If A has the same distribution as a degenerate aggregation of random variables, then it is degenerate.

Proof: For (1), consider A to be induced by $\hat{A}$. If I is not just one point, then the boundary conditions of $\hat{A}$ imply that there exist $\vec{x},\vec{y} \in I^n$ such that $\hat{A}(\vec{x}) \neq \hat{A}(\vec{y})$. Consider the random vector $\vec{X}$ such that $P(\vec{X} = \vec{x}) = P(\vec{X} = \vec{y}) = 0.5$. Then, it is clear that $A(\vec{X})$ has not degenerate distribution, thus A cannot be degenerate.

For (2), 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 $A(\vec{X}) =_{a.s.} \hat{A}(\vec{X},\vec{\lambda}_{\vec{X}})$ for any $\vec{X} \in L_I^n$ with $\hat{A} : I^n \times \mathbb{R}^d \to I$ and $(\vec{\lambda}_{\vec{X}},\vec{X} \in L_I^n)$ as in Theorem 3.22. If A is also conditionally determined, applying (3) in Theorem 3.18 one also has $A(\vec{X}) =_{a.s.} G_{\vec{X}}(\vec{X})$ for any $\vec{X} \in L_I^n$. Then, $\hat{A}(\vec{X},\vec{\lambda}_{\vec{X}}) =_{a.s} G_{\vec{X}}(\vec{X})$. The expectation $E[\hat{A}(\vec{x},\vec{\lambda}_{\vec{X}})]$, since equals $G_{\vec{X}}(\vec{x})$, always exists for any $\vec{x} \in I^n$ and $\vec{X} \in L_I^n$. In addition, since $\vec{\lambda}_{\vec{X}}$ has the same distribution for all $\vec{X} \in L_I^n$, $E[\hat{A}(\vec{x},\vec{\lambda}_{\vec{X}})]$ always takes the same value for a fixed $\vec{x} \in I^n$ and any $\vec{X} \in L_I^n$. Then, define $\hat{B} : I^n \to I$ as $\hat{B}(\vec{x}) = E[\hat{A}(\vec{x},\vec{\lambda}_{\vec{X}})]$. Since $\hat{A}$ is increasing in the first n components, $E[\hat{A}(\vec{x}_1,\vec{\lambda}_{\vec{X}})] \leq E[\hat{A}(\vec{x}_2,\vec{\lambda}_{\vec{X}})]$ if $\vec{x}_1 \leq \vec{x}_2$. If I has a lower bound a, then it is clear that $E[\hat{A}(a\vec{1},\vec{\lambda}_{\vec{X}})] = a$. If I does not have a lower bound, for any $x \in I$ there exists $\vec{x} \in I^n$ such that $\hat{A}(\vec{x},\vec{z}) < x$ for any $\vec{z} \in \mathbb{R}^d$. Then, $E[\hat{A}(\vec{x},\vec{\lambda}_{\vec{X}})] < x$. Proceeding similarly for the upper bound, it is concluded that $\hat{B}$ is increasing and fulfills the boundary conditions, thus is an aggregation function. It is concluded that

$A(\vec{X}) =_{a.s.} G_{\vec{X}}(\vec{X}) =_{a.s} \hat{B} \circ \vec{X}$ with $\hat{B}$ being an aggregation function. Therefore, it is almost surely equal to an induced aggregation of random variables.

The proof of (3) is equivalent as the latter one but replacing $=_{a.s.}$ by $=_{st}$.

For (4), consider for any $\vec{x} \in I^n$ the random vector $\vec{X}$ such that $P(\vec{X} = \vec{x}) = 1$. Then, since A has the same distribution as a conditionally determined aggregation of random variables, $A(\vec{X})$ should be degenerate. But, since it is randomly induced, one has that $A(\vec{X}) = \hat{A}(\vec{X}, \lambda_{\vec{X}})$. Recall that $\vec{\lambda}_{\vec{X}}$ has the same distribution for any $\vec{X} \in L_I^n$. If the distribution of $\lambda_{\vec{X}}$ is not degenerate, then there exists a function $\hat{B} : I^n \to I$ such that $\hat{B}(\vec{x}) = \hat{A}(\vec{x}, \vec{z})$ for any $\vec{z} \in \mathbb{R}^d$. If $\lambda_{\vec{X}}$ is degenerate with $P(\vec{\lambda}_{\vec{X}} = \vec{z}) = 1$, then define $\hat{B} : I^n \to I$ as $\hat{B}(\vec{x}) = \hat{A}(\vec{x}, \vec{z})$. In both cases, it is clear that $A(\vec{X}) = \hat{B}(\vec{X})$ and A is induced.

Finally, for (5) consider B a degenerate aggregation of random variables such that $A(\vec{X}) =_{st} B(\vec{X})$ for any $\vec{X} \in L_I^n$. If for any $\vec{X} \in L_I^n$ there exists $x \in I$ such that $P(B(\vec{X}) = x) = 1$, then it is straightforward that $P(A(\vec{X}) = x) = 1$ and, therefore, A is degenerate. ■

It is worth commenting that point (1) is an adaptation of Proposition 3.4. in [31], while the rest of the points can be found in [26]. In addition, point (4), with a slight modification, implies the non-trivial implication of point (2). For finishing to understand all the relationship between the families, 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. An example in this regard can be constructed using countermonotone random variables.

Example 3.28 *[26] Let $A : L_I^n \to L_I$ be an aggregation of random variables defined as $A(\vec{X}) = \max(\vec{X})$ for any $\vec{X}$ such that $\max(\vec{X})$ is not continuous and $A(\vec{X}) = Y$ with Y being a random variable that has the same distribution as $\max(\vec{X})$ and such that Y and $\max(\vec{X})$ are countermonotone if $\max(\vec{X})$ is continuous. Then, A has the same distribution as the induced random variable $\max(\vec{X})$, is conditionally determined since $A(\vec{X})$ can be expressed as a function of $\max(\vec{X})$, that is a function of $\vec{X}$, but is not induced.*

With all the relationships between the different families studied already, they are represented in Figure 3.1, in which each family is associated with a subset of

the plane and in which the intersection of these sets represents the intersection of the families.

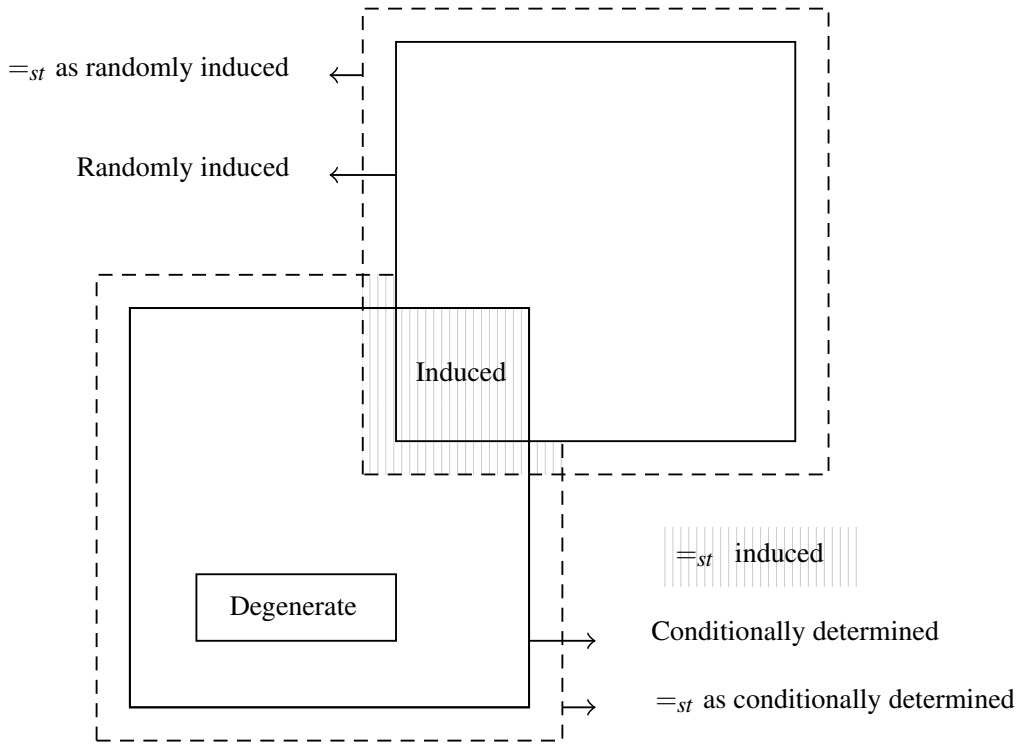

Figure 3.1: Representation of families of aggregations of random variables. The notation $=_{st}$ stands for "same distribution as" [26].

In addition, the study is also useful for identifying the different scenarios that aggregations of random variables can model. In particular, it has been seen that, in addition to aggregations with random inputs, it is possible to identify the random vector that is being aggregated in order to change the function that is applied, to change the dependence between the inputs and the output and to consider random parameters. In Figure 3.2, a representation of the resulting aggregations of random variables that appear when considering the different scenarios is provided.

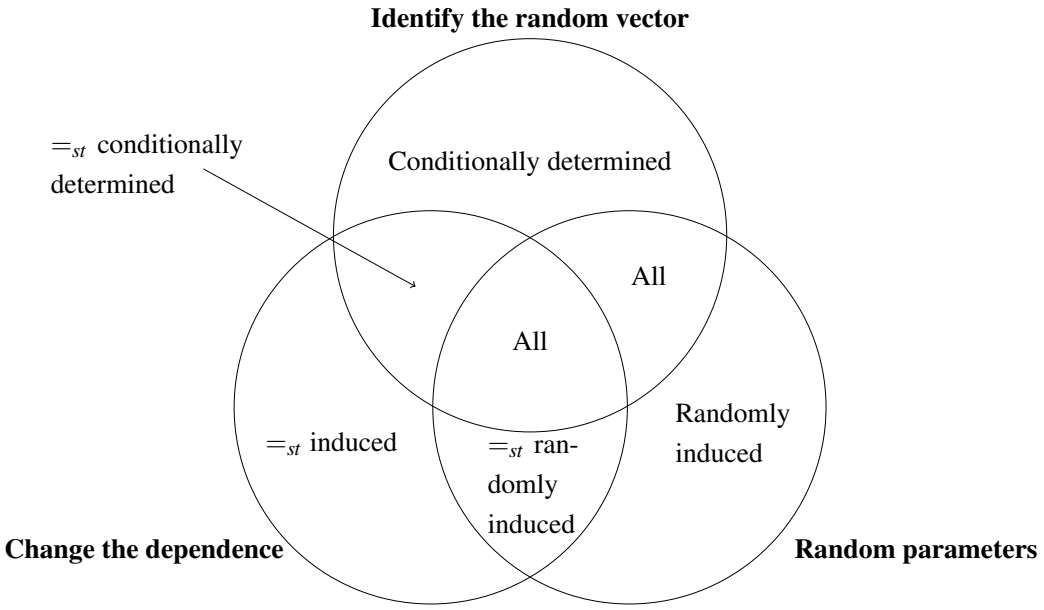

Figure 3.2: Families of aggregations of random variables associated with identifying the random vector, considering random parameters and changing the dependence. The notation $=_{st}$ stands for "same distribution as" [26].

3.3 Additional properties

In the last section, some properties, mainly related to probabilistic construction methods, have been used to define some families of aggregations of random variables. In this section, a different approach is considered, extending or studying the impact of well-known properties of aggregation functions to aggregations of random variables. In particular, many different definitions for idempotence and internality are given in Section 3.3.1, the consequences of associativity and recursivity in random series are studied in Section 3.3.2 and some extensions of monotonicity are provided in Section 3.3.3.

3.3.1 Idempotence and internality

Idempotence and internality are defined in terms of equalities and inequalities. When considering random variables, inequalities can be interpreted in terms of the

usual stochastic order, almost surely statements, conditional distributions or other perspectives such as working with expectations. In these sections, some concepts of idempotence and internality for aggregations of random variables are defined and their implications are studied.

3.3.1.1 Idempotence

The idea behind idempotence is "if the inputs are the same, the output is also the same". Although for real values the concept is clear, for random variables this could be interpreted in different ways. In the next definition, five different alternatives are considered.

Definition 3.29 *[33] Let I be a real interval and consider a function $f : L_I^n \to L_I$. Then,*

- *If $f(\vec{X}) =_{a.s.} X_1$ for any $\vec{X} \in L_I^n$ such that $X_1 =_{a.s.} \cdots =_{a.s.} X_n$, f is said to be almost surely idempotent,*
- *If $f(\vec{X}) =_{st} X_1$ for any $\vec{X} \in L_I^n$ such that $X_1 =_{st} \cdots =_{st} X_n$, f is said to be idempotent in distribution,*
- *If $P(f(\vec{X}) = \lambda \mid X_1 = \cdots = X_n = \lambda) = 1$ for any $\vec{X} \in L_I^n$ and $\lambda \in I$ such that $[f(\vec{X}) \mid X_1 = \cdots = X_n = \lambda]$ is well-defined, f is said to be conditionally idempotent,*
- *If $E[f(\vec{X}) \mid X_1 = \cdots = X_n = \lambda] = \lambda$ for any $\vec{X} \in L_I^n$ and $\lambda \in I$ such that $[f(\vec{X}) \mid X_1 = \cdots = X_n = \lambda]$ is well-defined, f is said to be martingale idempotent,*
- *If $P(f(\vec{X}) = \lambda) = 1$ for any $\vec{X} \in L_I^n$ such that $P(X_i = \lambda) = 1$ for any $i \in [n]$, f is said to be degenerate idempotent.*

Almost surely idempotence and idempotence in distribution work with equalities between random variables. On the other hand, conditionally idempotence and martingale idempotence focus on the values the random variables take. Finally,

degenerate idempotence is just the classical definition of idempotence restricted to degenerate random vectors.

Some notions of idempotence are stronger than others. In the next result, the implications between the definitions are stated. It should be noted that, at this point, the function $f : L_I^n \to L_I$ is a general one, not an aggregation of random variables. Therefore, in the next result monotonicity is not used.

Proposition 3.30 *[33] Let I be a real interval and consider a function $f : L_I^n \to L_I$. Then,*

1. *f is conditionally idempotent $\implies$ f is almost surely idempotent,*
2. *f is conditionally idempotent $\implies$ f is martingale idempotent,*
3. *f is almost surely idempotent $\implies$ f is degenerate idempotent,*
4. *f is idempotent in distribution $\implies$ f is degenerate idempotent.*

Proof:

1. If $X_1 =_{a.s.} \cdots =_{a.s.} X_n$, then $P(X_1 = \cdots = X_n) = 1$. Since for any $\lambda \in I$ $P(f(\vec{X}) = \lambda \mid X_1 = \cdots = X_n = \lambda) = 1$, then

   $$P(f(\vec{X}) = X_1 \mid X_1 = \cdots = X_n = \lambda) = P(f(\vec{X}) = X_1 \mid X_1 = \lambda) = 1,$$

   for any $\lambda \in I$ (if well defined). Applying Theorem 2.63, it is concluded that $f(\vec{X}) =_{a.s.} X_1$.

2. If $P(f(\vec{X}) = \lambda \mid X_1 = \cdots = X_n = \lambda) = 1$, then trivially $E[f(\vec{X}) \mid X_1 = \cdots = X_n = \lambda] = \lambda$.

3. If $P(X_i = \lambda) = 1$ for any $i \in [n]$, then $X_1 =_{a.s.} \cdots =_{a.s.} X_n$ so $f(\vec{X}) =_{a.s.} X_1$ and $P(f(\vec{X}) = \lambda) = 1$.

4. If $P(X_i = \lambda) = 1$ for any $i \in [n]$, then $X_1 =_{st} \cdots =_{st} X_n$ so $f(\vec{X}) =_{st} X_1$ and $P(f(\vec{X}) = \lambda) = 1$. ∎

It is concluded that the strongest properties are conditionally idempotence and idempotence in distribution. The weakest is the degenerate idempotence. Martingale idempotence, although it seems to be the least intuitive notion of idempotence, has an interesting implication related to the convex stochastic order.

Proposition 3.31 *[33] Let $A : L_I^n \to L_I$ be a martingale idempotent function. Then,*

$$[X_i \mid X_1 = \cdots = X_n] \leq_{cx} [A(\vec{X}) \mid X_1 = \cdots = X_n],$$

for any $i \in [n]$

Proof: Notice that

$$E\left[[A(\vec{X}) \mid X_1 = \cdots = X_n] \;\middle|\; [X_i \mid X_1 = \cdots = X_n] = \lambda\right] =$$
$$= E[A(\vec{X}) \mid X_1 = \cdots = X_n = \lambda] = \lambda.$$

The result holds by applying 2 in Theorem 2.117. ∎

Finally, it is possible to prove that aggregations of random variables induced by idempotent aggregations are conditionally idempotent.

Proposition 3.32 *[33] Let A be an aggregation of random variables induced by an idempotent aggregation function. Then, A is conditionally idempotent.*

Proof: Let $\hat{A}$ be the idempotent aggregation function that induces A. If $X_1 = \cdots = X_n = \lambda$, then $\hat{A}(\lambda, \ldots, \lambda) = \lambda$, thus $P(A(\vec{X}) = \lambda \mid X_1 = \cdots = X_n = \lambda) = 1$ whenever it is well-defined. ∎

3.3.1.2 Internality

This section is devoted to developing a similar study as the latter for internality. The idea behind being internal is "to be between the minimum and the maximum". Again, five different notions can be defined following this idea.

Definition 3.33 *[33] Let $f : L_I^n \to L_I$ be a function. Then:*

- *If* $\min(\vec{X}) \leq_{a.s.} f(\vec{X}) \leq_{a.s.} \max(\vec{X})$ *for any* $\vec{X} \in L_I^n$, f *is said to be almost surely internal,*
- *If* $\min(\vec{X}) \leq_{st} f(\vec{X}) \leq_{st} \max(\vec{X})$ *for any* $\vec{X} \in L_I^n$, f *is said to be internal in distribution,*
- *If* $\min_{i\in[n]} P(X_i \leq \lambda) \leq P(f(\vec{X}) \leq \lambda) \leq \max_{i\in[n]} P(X_i \leq \lambda)$ *for any* $\vec{X} \in L_I^n$ *and* $\lambda \in I$, f *is said to be internal in probability,*
- *If* $S(f(\vec{X})) \subseteq \left[\inf \cup_{i=1}^n S(X_i), \sup \cup_{i=1}^n S(X_i)\right]$, f *is said to be internal in support,*
- *If for any* $\vec{X} \in L_I^n$ *such that* $P(\vec{X} \in \vec{\lambda}) = 1$ *with* $\vec{\lambda} \in I^n$ *holds* $P(f(\vec{X}) \in [\min(\vec{\lambda}), \max(\vec{\lambda})]) = 1$, f *is said to be degenerate internal.*

For the first two properties, the maximum and minimum operators are applied to the random vectors and the aggregation is compared, using almost surely inequalities or the usual stochastic order, with them. The third and fourth alternatives take the maximum and minimum over distribution functions or the support of the random variables. Again, there is a notion for just the usual internality when working with degenerate distributions.

As in the other case, some implications between the introduced properties can be proved without using monotonicity.

Proposition 3.34 *[33] Let* $f : L_I^n \to L_I$ *be a function. Then:*

1. f *is almost surely internal* $\implies$ f *is internal in distribution,*
2. f *is internal in probability* $\implies$ f *is internal in distribution,*
3. f *is internal in distribution* $\implies$ f *is internal in support,*
4. f *is internal in support* $\implies$ f *is degenerate internal.*

Proof:

1. It is held trivially from the fact that $X \leq_{a.s.} Y \implies X \leq_{st} Y$.

2. Let $F(\vec{t})$, $F_{\min}(t)$ and $F_{\max}(t)$ be the distribution functions of $\vec{X}$, $\min(\vec{X})$ and $\max(\vec{X})$. In addition, denote as $F_1,\ldots,F_n$ the distribution functions of $X_1,\ldots,X_n$. Similarly, denote as $\bar{F}(\vec{t}),\bar{F}_{\min}(t),\bar{F}_{\max}(t),\bar{F}_1,\ldots,\bar{F}_n$ the associated survival functions.

   One has that $F_{\max}(t) = F(t\vec{1}) = C(F_1(t),\ldots,F_n(t))$, with C being a copula of $\vec{X}$. By the upper Fréchet Hoeffding bound, it is known that $C \leq \min$. In addition, one has that $\bar{F}_{\min}(t) = \bar{F}(t\vec{x}) = \bar{C}(\bar{F}_1(t),\ldots,\bar{F}_n(t))$, with $\bar{C}$ being a survival copula of $\vec{X}$. By the upper Fréchet Hoeffding bound, it is known that $\bar{C} \leq \min$. Therefore,

$$F_{\max}(\lambda) \leq \min_{i\in[n]} F_i(\lambda) = \min_{i\in[n]} P(X_i \leq \lambda)$$

   and,

$$F_{\min}(\lambda) = 1 - \bar{F}_{\min}(\lambda) \geq 1 - \min_{i\in[n]} \bar{F}_i(\lambda) = \max_{i\in[n]} P(X_i \leq \lambda).$$

   It is concluded that if

$$\min_{i\in[n]} P(X_i \leq \lambda) \leq P(f(\vec{X}) \leq \lambda) \leq \max_{i\in[n]} P(X_i \leq \lambda),$$

   then $F_{\max}(\lambda) \leq P(f(\vec{X}) \leq \lambda) \leq F_{\min}(\lambda)$, thus $\min(\vec{X}) \leq_{st} A(\vec{X}) \leq_{st} \max(\vec{X})$.

3. Recall that the support of any random variable is a closed set. Suppose that there exists $t \in \mathbb{R}$ such that $t < \inf U_{i=1}^n S(X_i)$ but $t \in S(f(\vec{X}))$. Then, there exists $\varepsilon > 0$ such that $P(f(\vec{X}) \in (t-\varepsilon,t+\varepsilon)) > 0$ but $P(X_i \in (t-\varepsilon,t+\varepsilon)) = 0$ for every $i \in [n]$. Moreover, it holds $F_{X_i}(t+\varepsilon) = 0$. Comparing the cumulative distribution functions of $f(\vec{X})$ and $\min(\vec{X})$:

$$F_{f(\vec{X})}(t+\varepsilon) \geq P(f(\vec{X}) \in (t-\varepsilon,t+\varepsilon)) > 0$$
$$F_{\min(\vec{X})}(t+\varepsilon) \leq \sum_{i=1}^{n} F_{X_i}(t+\varepsilon) = 0.$$

   It is concluded that $\min(\vec{X}) \not\leq_{st} f(\vec{X})$, which is a contradiction to the fact of f being internal in distribution. If there exists $t \in \mathbb{R}$ such that $t > \sup U_{i=1}^n S(X_i)$ but $t \in S(f(\vec{X}))$, the contradiction can be reached similarly.

4. Notice that if $P(X = \lambda) = 1$, $S(X) = \{\lambda\}$. ■

It is concluded that the strongest notions of internality are almost surely internality and internality in probability. It is easy to construct almost surely internal and internal in probability aggregations of random variables just by considering, respectively, induced aggregations of random variables by internal aggregation functions and distribution-based aggregations of random variables.

Proposition 3.35 *[33] Let $A : L_I^n \to L_I$ be an aggregation of random variables induced by an internal aggregation function. Then, A is almost surely internal.*

Proof: Let $\hat{A} : I^n \to I$ be the internal aggregation function that induces A. Since $\min(\vec{x}) \leq \hat{A}(\vec{x}) \leq \max(\vec{x})$ for any $\vec{x} \in I^n$, then $\min(\vec{X}) \leq_{a.s.} A(\vec{X}) \leq_{a.s.} \max(\vec{X})$ for any $\vec{X} \in L_I^n$. ■

Proposition 3.36 *[33] Let $A : L_I^n \to L_I$ be a distribution-based aggregation of random variables. Then, A is internal in probability.*

Proof: Let A be distribution-based on $\hat{A}$. For any $\vec{X} \in L_I^n$ with marginal distribution functions $F_1, \ldots, F_n$, $A(\vec{X})$ has as distribution function $\hat{A}(F_1, \ldots, F_n)$. Since $\hat{A}$ is internal, see Proposition 3.20, then $\min_{i\in[n]} F_i(t) \leq \hat{A}(F_1, \ldots, F_n)(t) \leq \max_{i\in[n]} F_i(t)$ for any $t \in I$, so A is internal in probability. ■

3.3.1.3 Relationship between idempotence and internality

For increasing functions, idempotence and internality are equivalent (see Proposition 2.10). In this section, the implications between the notions of idempotence and internality, when the function is an aggregation of random variables, are studied.

Proposition 3.37 *[33] Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then,*

1. *A is degenerate idempotent $\iff$ A is degenerate internal,*

2. *A is idempotent in distribution* $\implies$ *A is internal in distribution,*

3. *A is almost surely idempotent* $\implies$ *A is internal in distribution,*

4. *A is almost surely internal* $\implies$ *A almost surely idempotent,*

5. *A is internal in probability* $\implies$ *A is idempotent in distribution.*

Proof:

1. Let $\vec{X} \in L_I^n$ be a random vector such that $P(\vec{X} = \vec{\lambda}) = 1$ with $\vec{\lambda} \in I^n$. For the first implication, using monotonicity one has that, since $\min(\vec{\lambda})\vec{1} \leq_{st} \vec{X} \leq_{st} \min(\vec{\lambda})\vec{1}$, then $A(\min(\vec{\lambda})\vec{1}) = \min(\vec{\lambda}) \leq_{st} A(\vec{X}) \leq_{st} \max(\vec{\lambda}) = A(\max(\vec{\lambda})\vec{1})$. Therefore, $P(A(\vec{X}) \in [\min(\vec{\lambda}), \max(\vec{\lambda})]) = 1$. For the second implication, if $\vec{\lambda} = (\lambda, \ldots, \lambda)$, then $\min(\vec{\lambda}) = \lambda \leq_{st} A(\vec{X}) \leq_{st} \lambda = \max(\vec{\lambda})$, thus $P(A(\vec{X}) = \lambda) = 1$.

2. Starting from $\min(\vec{X})\vec{1} \leq_{st} \vec{X} \leq_{st} \max(\vec{X})\vec{1}$, as a consequence of monotonicity and idempotence in distribution it holds

$$\min(\vec{X}) =_{st} A(\min(\vec{X})\vec{1}) \leq_{st} A(\vec{X}) \leq_{st} A(\max(\vec{X})\vec{1}) =_{st} \max(\vec{X}).$$

3. Similarly as the previous case,

$$\min(\vec{X}) =_{a.s.} A(\min(\vec{X})\vec{1}) \leq_{st} A(\vec{X}) \leq_{st} A(\max(\vec{X})\vec{1}) =_{a.s.} \max(\vec{X}),$$

and the result follows since $X =_{a.s.} Y \implies X =_{st} Y$.

4. If $X_1 =_{a.s.} \cdots =_{a.s.} X_n$, then $\min(\vec{X}) =_{a.s.} \max(\vec{X}) =_{a.s.} X_1$. Since $\min(\vec{X}) \leq_{a.s} A(\vec{X}) \leq_{a.s.} \max(\vec{X})$, then $A(\vec{X}) =_{a.s.} X_1$.

5. Denote as $F_1, \ldots, F_n$ the distribution functions of $X_1, \ldots, X_n$. If $X_1 =_{st} \cdots =_{st} X_n$, then $F_1(t) = \cdots = F_n(t) = \min_{i\in[n]} F_i(t) = \max_{i\in[n]} F_i(t)$ for any $t \in \mathbb{R}$. Then,

$$F_1(t) = \min_{i\in[n]} F_i(t) \leq F_{A(\vec{X})}(t) \leq \max_{i\in[n]} F_i(t) = F_1(t),$$

for any $t \in \mathbb{R}$, thus $A(\vec{X}) =_{st} X_1$. ∎

As a consequence of this result, there is an equivalence between internality in support and degenerate internality for aggregations of random variables.

Corollary 3.38 *[33] Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then, A is internal in support if and only if A is degenerate internal.*

Proof: The first implication has been proved in Proposition 3.34. For the second one, using Proposition 3.37, any degenerate internal aggregation of random variables is degenerate idempotent. Consider $\vec{X} \in L_I^n$. If $\inf \cup_{i=1}^n S(X_i) = -\infty$ the condition over the infimum is fulfilled. Otherwise, suppose $\inf \cup_{i=1}^n S(X_i) = a > -\infty$. Using monotonicity, since $a\vec{1} \leq_{st} \vec{X}$, then $a =_{a.s.} A(a\vec{1}) \leq_{st} A(\vec{X})$ and $P(A(\vec{X}) \geq a) = 1$. Similarly, it is obtained that $P(A(\vec{X}) \leq b) = 1$ with $b = \sup \cup_{i=1}^n S(X_i)$. Therefore, $S(A(\vec{X})) \subseteq [a,b] = \left[\inf \cup_{i=1}^n S(X_i), \sup \cup_{i=1}^n S(X_i)\right]$. ∎

All the proved implications can be found in the diagram in Figure 3.3. The strongest properties are conditionally idempotence, almost surely internal and internal in probability. It is important to remark again that aggregations of random variables induced by idempotent/internal aggregation functions are conditionally idempotent and almost surely internal and distribution-based aggregations of random variables are internal in probability (Propositions 3.32, 3.35 and 3.36).

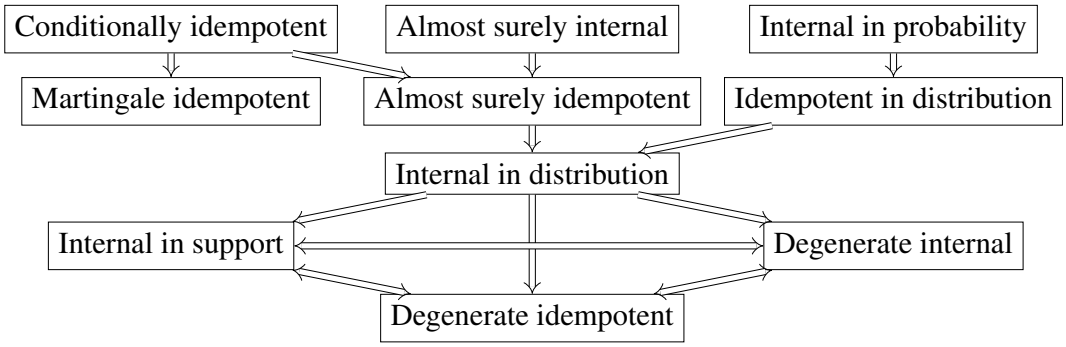

Figure 3.3: Implications between idempotence and internality notions for aggregations of random variables with respect to the usual stochastic order [33].

3.3.1.4 Counterexamples for the rest of implications

Propositions 3.30, 3.34 and 3.37 give some relations between the different properties. However, nothing has been said about the rest of possible implications. In the following, counterexamples for some of the implications that are not given in the diagram of Figure 3.3 can be found.

Example 3.39 *[33] Degenerate internality does not imply internality in distribution. Consider the degenerate aggregation of random variables* $A : L_{\mathbb{R}}^2 \to L_{\mathbb{R}}$ *defined by* $A(X_1,X_2) = \frac{1}{2}E[X_1+X_2]$. *It is degenerate internal. But if* X_1 *and* X_2 *are two standard uniform distributions, then*

$$\min_{i\in\{1,2\}} P(X_i \leq \lambda) = \max_{i\in\{1,2\}} P(X_i \leq \lambda) = \lambda.$$

Therefore, if A is internal in distribution, $A(X_1,X_2)$ *should have standard uniform distribution, which is not true since* $A(X_1,X_2) =_{a.s} \frac{1}{2}$.

Example 3.40 *[33] Martingale idempotence does not imply degenerate idempotence. Consider* $A : L_{\mathbb{R}}^2 \to L_{\mathbb{R}}$ *defined as* $A(X_1,X_2) = \frac{1}{2}(X_1+X_2+Y_{X_1,X_2})$, *where* Y_{X_1,X_2} *is a non-degenerate random variable with mean* 0 *and independent of* X_1 *and* X_2. *This aggregation of random variables is martingale idempotent but not conditionally idempotent, since if* $X_1 = X_2 = \lambda$, *then*

$$\left[\frac{1}{2}(X_1+X_2+Y_{X_1,X_2})) \;\middle|\; X_1 = X_2 = \lambda\right] = \lambda + \frac{Y_{X_1,X_2}}{2},$$

which has mean λ *but does not equal* λ *with probability one.*

Example 3.41 *[33] Conditional idempotence does not imply almost surely internality. Consider an interval I and, for each random vector* $\vec{X} \in L_I^2$, *a random variable* $Y_{\vec{X}}$ *independent of* $\vec{X}$ *and with the same distribution as* $\max(\vec{X})$. *Define the aggregation of random variables* $A : L_I^2 \to L_I$ *as,*

$$A(\vec{X}) = \begin{cases} \max(\vec{X}) & \text{if } X_1 = X_2, \\ Y_{\vec{X}} & \text{if } X_1 \neq X_2. \end{cases}$$

It is conditionally idempotent since

$$[A(\vec{X}) \mid X_1 = X_2 = \lambda] = \max(\lambda, \lambda) = \lambda.$$

However, if $X_1 =_{a.s.} 0$ *and* X_2 *has standard uniform distribution,* $\max(\vec{X}) = X_2$ *and* $A(X_1, X_2)$ *is independent of* X_2 *and holds* $A(X_1, X_2) =_{st} \max(X_1, X_2)$*, so* $A(X_1, X_2) =_{st} X_2$*. It is clear that* $P(A(X_1, X_2) > X_2) = 0.5$*. Then, A is not almost surely internal.*

Example 3.42 *[33] Neither conditionally idempotence nor almost surely internality implies idempotence in distribution. Consider the aggregation of random variables induced by the maximum. It is conditionally idempotent and almost surely internal by Propositions 3.32 and 3.35. Consider two random variables* X_1 *and* X_2 *with the same non-degenerate distribution and independent. Then, it is clear that* $\max(X_1, X_2) \neq_{st} X_1$.

Example 3.43 *[33] Internality in probability does not imply almost surely idempotence. Consider the aggregation of random variables distribution-based on the maximum and constructed using the standard uniform random variable* U*. It is internal in probability by applying Proposition 3.36. Consider* X_1 *a random variable independent of* U *and with non-degenerate distribution. One has* $A(X_1, X_1) =_{st} X_1$*, thus* $A(X_1, X_1)$ *has non-degenerate distribution and it is independent of* X_1*. It is concluded that* $A(X_1, X_1) \neq_{a.s.} X_1$.

The negation of the rest of the implications can be deduced from the latter ones and from Propositions 3.30, 3.34 and 3.37. For instance, suppose that internality in distribution implies almost surely idempotence. Then, since internality in probability implies internality in distribution, it should hold that internality in probability implies almost surely idempotence. This is not true, as has been shown in Example 3.43.

In this direction, it can be concluded that in Figure 3.3 all the implications between the different properties are represented. The implications that cannot be deduced from transitivity are false.

3.3.2 Associativity and recursivity

Associativity and recursivity are intrinsically related to the Markov property of stochastic processes. Consider $(X_n, n \in \mathbb{N})$ to be a sequence of independent random variables with the same distribution and independent. Then, given an associative extended aggregation function $A : \cup_{n\in\mathbb{N}} I^n \to I$, the sequence $(A^{(n)}(X_1,\ldots,X_n), n \in \mathbb{N})$ fulfills the Markov property.

Proposition 3.44 *[30] Let $(X_n, n \in \mathbb{N})$ be a sequence of independent random variables and let $A : \cup_{n\in\mathbb{N}} I^n \to I$ be an associative aggregation function. Then, the sequence $(A^{(n)}(X_1,\ldots,X_n), n \in \mathbb{N})$ is a Markov chain.*

Proof: Denote $Z_n = A^{(n)}(X_1,\ldots,X_n)$. Then, applying associativity, one has $Z_{n+1} = A^{(2)}(Z_n, X_{n+1})$. Given the value of Z_n, Z_{n+1} is a function of X_{n+1}. For any $k \leq n$, Z_k is a function of $X_1,\ldots,X_k$. Since $X_1,\ldots,X_k$ are independent of X_{n+1} and transformations of independent variables are also independent [278], it is concluded that Z_{n+1} and Z_k with $k \in [n-1]$ are conditionally independent given Z_n, that is, the Markov property holds. ∎

The last result also holds if recursivity [110] is considered. With this approach, annihilator elements can be seen as absorbent states.

Proposition 3.45 *[30] Let $(X_n, n \in \mathbb{N})$ be a sequence of independent random variables and let $A : \cup_{n\in\mathbb{N}} I^n \to I$ be an extended associative aggregation function. If a is an annihilator element of A, then a is an absorbent state of $(A^{(n)}(X_1,\ldots,X_n))_{n\in\mathbb{N}}$.*

Proof: Since $A^{(2)}(a,x) = a$, one has that the probability of $A^{(n)}(X_1,\ldots,X_n) = a$ given $A^{(n-1)}(X_1,\ldots,X_{n-1}) = a$ is 1. ∎

On the other hand, if the random variables have the same distribution, the neutral elements of the aggregation restart the evolution of the Markov chain.

Proposition 3.46 *[30] Let $(X_n, n \in \mathbb{N})$ be a sequence of independent random variables with the same distribution and let $A : \cup_{n\in\mathbb{N}} I^n \to I$ be an associative aggregation function. Then, if e is a neutral element of A,*

$$\left[A^{(n)}(X_{k+1},\ldots,X_{k+n}) \;\middle|\; A^{(n)}(X_1,\ldots,X_k) = e\right] =_{st} A^{(n)}(X_1,\ldots,X_n),$$

for any $k,n \in \mathbb{N}$.

Proof: Since $A^{(2)}(e,x) = x$, given $A^{(k)}(X_1,\ldots,X_k) = e$, $A^{(k+n)}(X_1,\ldots,X_{k+n}) = A^{(n)}(X_{k+1},\ldots,X_{k+n})$. The random variables in $(X_n, n \in \mathbb{N})$ have the same distribution, thus $A^{(n)}(X_{k+1},\ldots,X_{k+n})$ has the same distribution as $A^{(n)}(X_1,\ldots,X_n)$. ∎

This equivalence between associative extended aggregation functions on some particular type of random sequences and Markov chains will be especially relevant in Sections 5.1.3 and 5.1.4 , in the particular case of idempotent nullnorms and uninorms.

3.3.3 Another notions of monotonicity

Some of the non-intuitive behaviors of aggregations of random variables are possible because monotonicity only takes into account the distribution of the input and the output, not the values that they take for different values of Ω or their dependence. In this section, alternative notions of monotonicity are presented. They are focused on other approaches rather than on the change in the used stochastic order, since this has already been studied in Section 3.1.2.

The first one will be conditional monotonicity. For conditionally determined aggregation functions, it is possible that, when the value of the input random vector increases, the value of the output random variable decreases. For instance, this happens for some random vectors in the aggregation of random variables given in Example 3.28. In this direction, conditional monotonicity is defined in order to impose a positive relation between the values that the input random vector and the output random variable take.

Definition 3.47 *[39] Let* $A : L_I^n \to L_I$ *be a conditionally determined aggregation of random variables. Then, A is said to be conditionally monotone if* $[A(\vec{X}) \,|\vec{X} = \vec{x}] \leq_{st} [A(\vec{X}) \,|\vec{X} = \vec{y}]$ *for any* $\vec{X} \in L_I^n$ *and* $\vec{x},\vec{y} \in I^n$ *such that* $\vec{x} \leq \vec{y}$ *and the latter conditional distributions are well-defined.*

Notice that $\leq_{a.s.}$ could be written instead of $\leq_{st}$, since both $[A(\vec{X}) \,|\vec{X} = \vec{x}]$ and $[A(\vec{X}) \,|\vec{X} = \vec{y}]$ have degenerate distribution. This property can be seen as a sort

of comonotonicity between the initial random vector and the output random variable. Another interpretation is that conditional monotone aggregations applied over observations of the random vectors are increasing functions from vectors to real numbers. This will be especially relevant for prediction problems, in which monotonicity regarding the input values and the prediction outcome is usually needed. If the aggregation of random variables is not conditionally determined, the dependence can also be considered as follows.

Definition 3.48 *Let $A : L_I^n \to L_I$ be an aggregation of random variables. Then, A is said to be jointly monotone if $(\vec{X},A(\vec{X})) \leq_{st} (\vec{Y},A(\vec{Y}))$ for any $\vec{X},\vec{Y} \in L_I^n$ such that $\vec{X} \leq \vec{Y}$.*

Notice that, using monotonicity, one has $\vec{X} \leq_{st} \vec{Y}$ and $A(\vec{X}) \leq_{st} A(\vec{Y})$, but they do not imply $(\vec{X},A(\vec{X})) \leq_{st} (\vec{Y},A(\vec{Y}))$, since the dependence between the input random vector and the output random variables should be taken into account. It can be proved that induced aggregations fulfill both properties.

Proposition 3.49 *Let $A : L_I^n \to L_I$ be an induced aggregation of random variables. Then, A is conditionally and jointly monotone.*

Proof: Since A is induced, it is conditionally determined. Let $\hat{A} : I^n \to I$ be the aggregation function that induces A. Then, using the monotonicity of $\hat{A}$,

$$[A(\vec{X}) \mid \vec{X} = \vec{x}] = \hat{A}(\vec{x}) \leq \hat{A}(\vec{y}) = [A(\vec{X}) \mid \vec{X} = \vec{y}],$$

for any $\vec{X} \in L_I^n$ and $\vec{x},\vec{y} \in I^n$ such that $\vec{x} \leq \vec{y}$ and the latter conditional distributions are well-defined. Therefore, A is conditionally monotone.

If $\vec{X} \leq_{st} \vec{Y}$, applying 3 in Theorem 2.102, consider $\hat{\vec{X}}$ and $\hat{\vec{Y}}$ such that $\hat{\vec{X}} =_{st} \vec{X}$, $\hat{\vec{Y}} =_{st} \vec{Y}$ and $\hat{\vec{X}} \leq_{a.s.} \hat{\vec{Y}}$. Then, using the monotonicity of $\hat{A}$, it is clear that

$$(\vec{X},A(\vec{X})) =_{st} (\hat{\vec{X}},\hat{A}(\hat{\vec{X}})) \leq_{a.s.} (\hat{\vec{Y}},\hat{A}(\hat{\vec{Y}})) =_{st} (\vec{Y},A(\vec{Y})),$$

for any $\vec{X},\vec{Y} \in L_I^n$ such that $\vec{X} \leq_{st} \vec{Y}$. ∎

Notice that there are examples of aggregations of random variables that are conditionally monotone or jointly monotone but are not induced. For instance, the

aggregations given in Example 3.11 and Example 3.23 are, respectively, conditionally monotone and jointly monotone, but none of them are induced.

Moving to weaker notions of monotonicity, a similar concept as directionally monotone pre-aggregations can be defined for the aggregation or random variables.

Definition 3.50 *Let (Ω, Σ, P) be a probability space and I a real non-empty interval. A pre-aggregation function of random variables (with respect to $\leq_{st}$) is a function $A : L_I^n(\Omega) \to L_I(\Omega)$ which satisfies:*

1. *There exists $\vec{r} \in \mathbb{R}^n$ such that $A(\vec{X}) \leq_{st} A(\vec{X} + c\vec{r})$ for any $\vec{X} \in L_I^n$ and $c \in \mathbb{R}^+$ such that $\vec{X} + c\vec{r} \in L_I^n$,*
2. *For any $X \in L_I$, there exists $\vec{X} \in I$ such that $A(\vec{X}) \leq_{st} X$,*
3. *For any $X \in L_I$, there exists $\vec{X} \in L_I^n$ such that $A(\vec{X}) \geq_{st} X$.*

Similarly as in the case of aggregation functions, it is possible to define a notion of induced pre-aggregation function by proving a Composition Theorem for them.

Theorem 3.51 *Let $\hat{A} : I^n \to I$ be a measurable pre-aggregation function. Then, the function $A : L_I^n \to L_I$ defined as $A(\vec{X}) = \hat{A} \circ \vec{X} = \hat{A}(\vec{X})$ is a pre-aggregation of random variables.*

Proof: Notice that, for any $\vec{X} \in L_I^n$, $\vec{X} : \Omega \to I^n$ and, since $\hat{A} : I^n \to I$, then $\hat{A}(\vec{X}) : \Omega \to I$. The measurability of $\hat{A}(\vec{X})$ is a consequence of the measurability of both $\vec{X}$ and $\hat{A}$. Therefore, $\hat{A}(\vec{X}) \in L_I$ and A is well-defined.

For the directional monotonicity, there exists $\vec{r} \in \mathbb{R}^n$ such that $\hat{A}(\vec{x}) \leq \hat{A}(\vec{x} + c\vec{r})$ for any $\vec{x} \in I$ and $c \in \mathbb{R}^+$ such that $\vec{x} + c\vec{r} \in I^n$. It is clear that for any $\vec{X} \in L_I^n$ and $c \in \mathbb{R}^+$ such that $\vec{X} + c\vec{r} \in L_I^n$, $\hat{A}(\vec{X}) \leq_{a.s.} \hat{A}(\vec{X} + c\vec{r})$ and, therefore, $A(\vec{X}) \leq_{st} A(\vec{X} + c\vec{r})$. For the boundary conditions, the proof is equivalent to the one in Theorem 3.12 ∎

The Induced Ordered Linear Fusion (IOLF) operator will be a particular case of a pre-aggregation of random variables studied in Section 4.1.

Chapter 4

Prominent aggregations of random variables

Contents

In the last chapter, a general theory of aggregations of random variables has been presented. The theory proposes a flexible definition that fits in many different scenarios. In addition, many properties, defined with an abstract approach, have been provided. This chapter is devoted to define and study particular aggregations of random variables, related functions and properties that are relevant in topics such as time series forecasting, prediction, estimation or random generation. Although the approach and field of application change drastically for each of the cases, they are all contained in the theory of aggregation of random variables previously developed.

Firstly, the Induced Ordered Linear Fusion (IOLF) operator will be introduced as a generalization of the use of the IOWA operator in time series forecasting (Section 4.1). Then, some techniques used in Aggregation Theory are applied to study

the asymptotic behavior of the optimal weights of L-statistics for mean estimation (Section 4.2). Later on, uniform random capacities are considered, as a way to study random generators of capacities. In addition, they allow one to define randomly induced versions of the Choquet and Sugeno integrals (Section 4.3). Subsequently, Stochastically Ordered Aggregation operators, which are conditionally determined aggregations of random variables (with respect to $\leq_{sd-st}$) that order the inputs by considering their distribution but not their particular values, are defined and studied (Section 4.4). Finally, the approximation of a random variable by means of minimizing the expected value of the absolute difference when a copula is fixed and the monotonicity of such a problem is disclosed (Section 4.5).

4.1 The Induced Ordered Linear Fusion operator

In time series forecasting, there are many methods and techniques used to make predictions. In order to obtain more robust and precise predictions, sometimes more than one model is trained and the predictions are fused using an ensemble. In many cases, the ensemble is an aggregation function.

This section is devoted to defining the IOLF operator, a generalization of the IOWA operator that deals with negative weights and that will be of interest as an ensemble for time series predictions. Firstly, a literature review on the use of aggregation functions in time series is provided and the usual approach in the use of the IOWA operator in this regard is explained. Secondly, the operator is defined and its properties are studied, focusing on its use as a mean estimator when a sample of non-identically distributed and possibly dependent random variables is given.

This operator will be used in an applied problem related to temperature and humidity prediction in Section 7.6.

4.1.1 Aggregation functions in time series forecasting

The main motivation for the definition of the IOLF operator is to generalize the IOWA operator as an ensemble method in time series forecasting. In the following, a survey on the topic of aggregation functions as ensemble methods for time series is

provided. The literature review has been carried out through the search engines *Scopus, Web of Science* and *Google Scholar*. The most important keywords used have been *time series, prediction, forecast, ensemble, aggregation, fusion, weighted averaging, ordered weighted averaging, induced weighted averaging, OWA and IOWA*. It has been used forward (snowball) and backward reference searching, considering papers that cite or are cited by relevant works.

The use of ensembles in time series has been deeply investigated in the literature. Some surveys focused on the topic have been published in recent years, for instance, see [5, 284, 320]. Technically, models based on decision trees such as random forests and boosted decision trees are particular types of ensembles (see [249]). However, the initial prediction models are weak predictors of the same type, while when aggregation functions are used, strong prediction models based on different techniques are usually considered.

For instance, the authors of [158] combine machine learning models using weighted arithmetic means in an application regarding financial data. The different architectures of the models make the ensembled model more flexible and, therefore, to have better accuracy than the initial models. Aggregation functions are widely used as ensembles in the literature [5]. For instance, the authors in [172] give a method that constructs confidence intervals based on aggregation functions. A comparative study between different alternatives of linear aggregation functions is made in [206]. In the considered datasets, the Ordered Weighted Averaging outperforms the weighted arithmetic mean.

Another aggregation function that is used in this regard is the Induced Ordered Weighted Averaging (IOWA), in which the predictions are ordered by, typically, the precision in the previous times. This technique is quite popular, with applications in safety monitoring [61, 309], logistic [210], economy [154, 174, 207, 337], energy resources [208, 215, 310] or climatology [209, 338]. The approach permits to rearrange the prediction models dynamically as time changes. Another alternative of a time-dependent ensemble is explored in [86], in which a weighted arithmetic mean with dynamic weights is proposed.

Other ensemble methods focus on how the initial prediction models are trained. Bagging ensemble (see [5]), uses different bootstrap samples to train the prediction models and then applies a weighted arithmetic mean. Note that bootstrap sampling

in time series is not easy due to the dependence between observations. Thus, block bootstrap [204] should be used. In these models, the diversity of the predictors is due not only to differences in the architecture but also to differences in the training samples.

Not all ensembles are aggregation functions. The predictions of the initial models can be used as inputs of a machine learning model that gives the final prediction. The authors in [273] use, for instance, Support Vector Regression. In [271], a Combiner is proved to have better behavior than different weighted-based aggregation functions for medical data. However, these ensemble models lack the explainability of weighted-based aggregation functions, as pointed out in [283].

In Table 4.1 a brief summary of some of the latter papers is presented, focusing on the used ensemble methods, the conclusions of the research and some relevant comments. Notice that WAM, OWA, IOWA and SVR stand for, respectively, Weighted Arithmetic Mean, Ordered Weighted Averaging, Induced Ordered Weighted Averaging and Support Vector Regression.

4.1.2 The usual procedure of IOWA in time series forecasting

In this section, the usual procedure for the IOWA operator applied to time series forecasting is briefly explained. Let $x_1,\dots,x_T$ be a finite sequence of numbers. In addition, let $p_1,\dots,p_n$ be n different prediction models for the latter sequence. For each $t \in \{2,\dots,T\}$, they give the predictions $\vec{p}_t = (p_{1,t},\dots,p_{n,t})$. The fused prediction p_t by the IOWA operator is defined as:

$$p_t = \text{IOWA}(\vec{p}_t, \vec{y}_t; \vec{w}), \tag{4.1}$$

where $\vec{w}$ is a weighting vector and $\vec{y}_t = (y_{1,t},\dots,y_{n,t})$ is defined as the precision of the prediction in the last time step,

$$y_{i,t} = \begin{cases} 1 - \left|\frac{x_{t-1}-p_{i,t-1}}{x_{t-1}}\right| & \text{if } \left|\frac{x_{t-1}-p_{i,t-1}}{x_{t-1}}\right| < 1, \\ 0 & \text{if } \left|\frac{x_{t-1}-p_{i,t-1}}{x_{t-1}}\right| \geq 1. \end{cases} \tag{4.2}$$

This approach is similar to the best yesterday's model introduced in [328], where the predictions of one day are ordered using the absolute difference between the last prediction and the observed value, for all the prediction models.

Ref	Ensemble model/s	Comments/Conclusions
[61]	WAM, IOWA	IOWA outperforms WAM
[206]	WAM, OWA	OWA outperforms WAM
[158]	WAM	Ensemble outperforms initial models
[172]	Based on the Mean	Focuses on confidence intervals
[309]	IOWA	Ensemble outperforms initial models
[210]	IOWA	Learning the weights improves the results
[207]	IOWA	Ensemble outperforms initial models
[154]	WAM, IOWA	Accuracy-based WAM has better results
[337]	IOWA	Ensemble outperforms initial models
[208]	IOWA	Ensemble outperforms initial models
[310]	IOWA	Ensemble outperforms initial models
[215]	Based on IOWA	Ensembles outperform initial models
[209]	IOWA	Ensemble outperforms initial models
[338]	IOWA	Uses bootstrap
[273]	SVR	SVR ensemble outperforms other models
[271]	Combiner, others	Combiner outperforms other ensembles
[283]	WAM, Median, Mode	Mode outperforms other ensembles

Table 4.1: Table summarizing the ensemble models used and the relevant conclusions or comments from some of the reviewed articles [35].

Linking this problem to the classical estimation of random samples, it can be seen as a more general case in which the random variables are not necessarily independent and with the same distribution. A reasonable assumption is to consider the sequence of observations as a random sequence $(X_1,\dots,X_T)$ and the prediction models, for each time $t \in \{2,\dots,T\}$, as a random vector $\vec{P}_t = (P_{1,t},\dots,P_{n,t})$. Another intuitive condition is to consider that, on average, the prediction models give the correct value of the time series. That is, $E[P_{i,t} \mid X_t] = X_t$ for any $i \in [n]$. In addition, for each time $t \in \{2,\dots,T\}$ there is also an additional random vector $\vec{Y}_t$, which is used to determine the permutation of the input values in the IOWA operator.

In order to construct the IOWA operator, one has to determine the weights. The classical way to optimize such a choice is to solve the following optimization

problem:

$$\begin{aligned} &\text{Minimize } \sum_{t=1}^{T}\left(x_t - \sum_{i=1}^{n} w_i \pi_{\vec{y}_t}(\vec{x}_t)_i\right)^2, \\ &\text{Subject to } \sum_{i=1}^{n} w_i = 1 \text{ and } w_1,\ldots,w_n \geq 0. \end{aligned} \tag{4.3}$$

which minimizes the Mean Squared Error in the considered sample. This problem is similar to the one presented in Theorem 4.6 when using the sample covariance matrix and the sample mean as estimators, but removing the positivity of the weights.

4.1.3 Definition and properties

The optimization problem defined in Equation 4.3 can be modified by not imposing positive values on the weights. By doing that, the feasible region in the optimization problem is bigger and a closed expression for its solution (see Theorem 4.6) can be computed. This consideration, which might be interesting for some applications, leads to an extension of the IOWA operator that will be called the Induced Ordered Linear Fusion.

Definition 4.1 *[35] Consider a generalized weighting vector $\vec{w} \in \mathbb{R}^n$ such that $\sum_{i=1}^{n} w_i = 1$ and the permutation $\pi_{\vec{y}} : [n] \to [n]$ such that $\pi_{\vec{y}}(\vec{y})_i = y_{(i)}$ and, if there is any draw in $\vec{y}$, replace the associated values of $\vec{x}$ by their average. Then, the Induced Ordered Linear Fusion (IOLF) has the following expression:*

$$IOLF(\vec{x},\vec{y};\vec{w}) = \vec{w}^t \pi_{\vec{y}}(\vec{x}),$$

Trivially, any IOWA is an IOLF operator.

4.1.3.1 Initial properties and semantic of negative weights

Negative weights, although they are a simple way to extend the IOWA operator, increase substantially the complexity of properties, interpretation and applicability of the resulting function. Firstly, many of the properties related to being an aggregation function, monotonicity and the boundary conditions, are not longer true. Therefore,

it is necessary to work with weaker alternatives of these properties. Secondly, the semantics of negative weights are not intuitive and should be studied. Finally, negative weights allow to capture more complex dependence structures between the aggregated values. These three points should be discussed in detail.

Starting with the properties, ratio scale invariance and idempotence remain to hold for the IOLF, since the condition $\sum_{i=1}^{n} w_i = 1$ is not changed. Therefore, the IOLF operator is weakly monotone. Moreover, following the same procedure as in the case of the OWA operators with negative weights (see Proposition 2.31), the IOLF operator with weights $\vec{w} \in \mathbb{R}^n$ is directionally increasing with respect any positive vector $\vec{r} \in \mathbb{R}^n$ such that:

$$\sum_{i=1}^{n} r_{\pi(i)} w_i > 0,$$

for any possible permutation π. Therefore, if $I = \mathbb{R}$, the IOLF operator is a pre-aggregation function.

However, if the interval is not the real line, the negative weights do not allow the IOLF operator to be a pre-aggregation function. For instance, if $I = [0,1]$ and $\vec{w} = (-0.5, 1.5)$, then:

$$\text{IOLF}((1,1),(0,2);(-0.5,1.5)) = -0.5 \notin [0,1].$$

Nevertheless, focusing on the main purpose, when predicting values of time series, the prediction models typically do not give predictions on a bounded interval but the real line [213]. For the cases in which it is mandatory to have the result in a specific interval, it is possible to consider a function from the real line to the interval. In the previous case, one can transform the predictions into the real line using the inverse of the sigmoid function [62] and apply the procedure there.

However, even considering $\mathbb{R}$ as the interval, having negative weights does not allow the IOLF to be monotone, making it impossible to be an aggregation function. The non-monotonicity may be unintuitive, since if the prediction of one of the prediction models increases, it is expected that the fused prediction will also increase. However, as will be seen later, this proposal could be better from a statistical point of view, even in the case the monotonicity is not preserved. Allowing negative weights expands the feasible region in the problem stated in Equation 4.3.

Therefore, a better optimal solution is expected. Also notice that pre-aggregations, even with the loss of monotonicity, are starting to be used in applied problems (see [112, 220, 221, 308]). In Section 7.6, numerical results will show that the best alternative for the considered datasets is a weighted arithmetic mean with negative weights, outperforming the rest of classical alternatives.

Moving to interpretability, there are clear difficulties in having a good semantic of negative weights. For the IOWA operator, the value of w_i can be interpreted as the importance of the prediction of the model with the i-th best prediction in the previous time step. For the OWA and WAM operators, the value of w_i is related to the importance of the i-th model or the i-th smallest prediction. Although the loss of semantics when using negative weights is evident, it is possible to interpret the absolute value of w_i as the importance of the prediction of the model associated with the i-th component of the ordered vector, considering also as important models the ones with a bad expected prediction.

From the point of view of probability, the optimal weights that will be obtained in Theorem 4.6 contain negative values if there is a strong positive partial correlation, see the paragraph right after the result. A strong positive dependence has consequences in terms of monotonicity. If just the prediction with an associated negative weight is considered, it is true that the aggregation decreases if the prediction increases. However, since a negative optimal weight implies a positive dependence, one can expect that an increase in the prediction implies an increase in the rest of the predictions, and therefore, an increase in the final aggregation. Following the interpretation in terms of covariance matrices, negative weights appear as a way to fine-tune co-increasing sets of predictions. In general, one can expect this behavior since, if all prediction models are adequate, they should increase and decrease together.

4.1.3.2 Probabilistic properties

In the following, it will be assumed that the predictions of the different models can be modeled as random variables. Of course, before moving on to more involved results, it is necessary to prove that the IOLF operator applied to random vectors is itself a random variable. Recall that a generalized weighting vector is any real

vector $\vec{w}$ such that $\sum_{i=1}^{n} w_i = 1$.

Proposition 4.2 *[35] Let $\vec{X}$ and $\vec{Y}$ be two random vectors and let $\vec{w} \in \mathbb{R}^n$ be a generalized weighting vector. Then, the function IOLF$(\vec{X}, \vec{Y}; \vec{w})$ is a random variable.*

Proof: Firstly, prove that $\pi_{\vec{Y}}(\vec{X})$ is a random vector. Consider the probability space $(\Omega, \mathscr{F}, P)$ on which $\vec{X}$ and $\vec{Y}$ are defined. It suffices to prove that any component is a measurable function from Ω to $\mathbb{R}$. In particular, it is enough to prove that $\pi_{\vec{Y}}(\vec{X})_i^{-1}((\infty, a]) \in \mathscr{F}$ for any $a \in \mathbb{R}$ and $i \in [n]$ [66].

Consider the case in which $\pi_{\vec{Y}}(\vec{Y})_i = Y_k$. Then, $\pi_{\vec{Y}}(\vec{X})_i^{-1}((\infty, a]) = X_k^{-1}((\infty, a])$. Since $\vec{X}$ is a random vector, this set belongs to $\mathscr{F}$.

This only happens if $Y_{j_1} < \cdots < Y_{j_i} < \cdots < Y_{j_n}$ with $(j_1, \ldots, j_n)$ being a permutation of $[n]$ such that $j_i = k$. Consider the following set,

$$B_k^i = \{\omega \in \Omega \mid \exists \pi : [n] \to [n] \text{ with } \pi(i) = k \text{ and } Y_{\pi(1)} \leq \cdots \leq Y_{\pi(n)}\},$$

which belongs to $\mathscr{F}$ by applying recursively that the intersection of measurable sets is measurable and that the set $\{\omega \in \Omega \mid X(\omega) \leq Y(\omega)\}$ is measurable for any pair of random variables X and Y [93].

Then, considering all the possible cases:

$$\pi_{\vec{Y}}(\vec{X})_i^{-1}((\infty, a]) = \cup_{k=1}^{n} \left(X_k^{-1}((\infty, a]) \cap B_k^i\right).$$

Since it is the (finite) union of measurable sets, then $\pi_{\vec{Y}}(\vec{X})_i^{-1}((\infty, a]) \in \mathscr{F}$ for any $i \in [n]$ and $\pi_{\vec{Y}}(\vec{X})$ is a random vector. Thus, any linear combination of their components is a random variable, and in particular it holds for IOLF$(\vec{X}, \vec{Y}; \vec{w})$. ∎

As a direct consequence of the latter result and Theorem 3.51, it is concluded that the IOLF operator, when the vector $\vec{y}$ is fixed, is a pre-aggregation of random variables. In the following, two extreme cases in which the expected value with respect to $\vec{Y}$ of the IOLF operator has a simple expression are provided.

Proposition 4.3 *[35] Let $\vec{X}$ and $\vec{Y}$ be two random vectors and $\vec{w} \in \mathbb{R}^n$ a generalized weighting vector. The following statements hold:*

1. *If $\vec{Y}$ and $\vec{X}$ are independent, then:*

$$E\left[IOLF(\vec{X},\vec{Y};\vec{w}) \;\middle|\; \vec{X}\right] = \sum_{k=1}^{n} \hat{w}_k X_k,$$
$$\hat{w}_k = \sum_{i=1}^{n} w_i P(Y_k = Y_{(i)}).$$

2. *If $\vec{Y}$ and $\vec{X}$ are independent and $\vec{Y}$ is exchangeable, then*

$$E\left[IOLF(\vec{X},\vec{Y};\vec{w}) \;\middle|\; \vec{X}\right] = \sum_{k=1}^{n} \frac{1}{n} X_k.$$

Proof: If $\vec{X}$ and $\vec{Y}$ are independent, the distribution of $[X_k \mid Y_k = Y_{(i)}]$ is the same as the distribution of X_i for any $i,k \in [n]$. Thus, the distribution function of a component of $\pi_{\vec{Y}}(\vec{X})_i$ can be expressed as

$$F_{\pi_{\vec{Y}}(\vec{X})_i} = \sum_{k=1}^{n} P\left(Y_k = Y_{(i)}\right) F_{X_k \mid Y_k = Y_{(i)}} = \sum_{k=1}^{n} P\left(Y_k = Y_{(i)}\right) F_{X_k},$$

for any $i \in [n]$. Then, for a fixed value of $\vec{X}$, computing the expectation leads to the following expression:

$$E\left[\mathrm{IOLF}(\vec{X},\vec{Y};\vec{w}) \;\middle|\; \vec{X}\right] = E\left[\sum_{i=1}^{n} w_i \pi_{\vec{Y}}(\vec{X})_i \;\middle|\; \vec{X}\right] =$$
$$= \sum_{i=1}^{n} w_i E\left[\pi_{\vec{Y}}(\vec{X})_i \;\middle|\; \vec{X}\right] = \sum_{i,k=1}^{n} w_i P\left(Y_k = Y_{(i)}\right) X_k = \sum_{k=1}^{n} \hat{w}_k X_k.$$

For the second point, notice that if $\vec{Y}$ is exchangeable, then $P\left(Y_k = Y_{(i)}\right) = \frac{1}{n}$ for any $i,k \in [n]$. ∎

It is important to remark that the most interesting cases are the intermediate ones, in which the order induced by $\vec{Y}$ is not the same as the order induced by $\vec{X}$ and neither $\vec{X}$ and $\vec{Y}$ are independent.

The mean vector and the covariance matrix of $\mathrm{IOLF}(\vec{X},\vec{Y};\vec{w})$, in terms of the random vector $\pi_{\vec{Y}}(\vec{X})$, have easy expressions based on the properties of the mean and variance of a linear transformation of a random vector. Their proof is straighforward and thus omitted.

Proposition 4.4 *[35] Consider two random vectors $\vec{X}$ and $\vec{Y}$ and let $\vec{w} \in \mathbb{R}^n$ be a generalized weighting vector. Then, $IOLF(\vec{X},\vec{Y};\vec{w})$ satisfies*

$$E\left[IOLF(\vec{X},\vec{Y};\vec{w})\right] = \vec{w}^t\vec{\mu},$$
$$Var\left[IOLF(\vec{X},\vec{Y};\vec{w})\right] = \vec{w}^t\Sigma\vec{w},$$

where $\vec{\mu}$ and Σ denote, respectively, the mean vector and the covariance matrix of $\pi_{\vec{Y}}(\vec{X})$.

4.1.3.3 Mean estimation using the IOLF operator

In this section, the applicability of the IOLF operator over random variables in the estimation of a common mean is studied. Notice that the main motivation is that the random variables will be associated with the predictions of several time series forecasting models and the common mean can be identified as the real value of the time series at that time.

In this direction, consider a random vector $\vec{X}$ for which the mean of every component is equal to the same value μ. Decompose $\vec{X} = \mu\vec{1} + \vec{Z}$, being $\vec{Z}$ a noise random vector for which the mean of all components is 0 and μ the parameter to be estimated. In addition, it will be considered that $\vec{Y}$ does not depend on μ, because in this case it might be tempting to use $\vec{Y}$ to estimate it. Finally, it is assumed that the covariance matrix of $\pi_{\vec{Y}}(\vec{X})$ is invertible. No additional considerations are made for $\vec{X}$ and $\vec{Y}$, the involved random variables can be dependent and may have different distributions. This is a very flexible scheme in which the prediction models can have a quite different behavior and can be correlated. In the simplest case, they may be a collection of $2n$ independent and identically distributed random variables.

In the rest of this section, it will be denoted $E\left[\pi_{\vec{Y}}(\vec{X})\right] - \mu\vec{1}$ as $\vec{\Delta}$, i.e. the drift with respect to the common mean. In addition, the covariance matrix of $\pi_{\vec{Y}}(\vec{X})$ will be denoted as Σ. With this change of notation and noticing that the elements of $\vec{w}$ sum 1, the expressions of Proposition 4.4 have the following form:

$$E\left[\text{IOLF}(\vec{X},\vec{Y};\vec{w})\right] = \mu + \vec{\Delta}^t\vec{w},$$
$$\text{Var}\left[\text{IOLF}(\vec{X},\vec{Y};\vec{w})\right] = \vec{w}^t\Sigma\vec{w}.$$

Notice that, since $\sum_{i=1}^{n} \pi_{\vec{Y}}(\vec{X}) = \sum_{i=1}^{n} X_i$, then it holds that $\sum_{i=1}^{n} \Delta_i = 0$. In addition, it is clear that $E\left[\left(\mu - \text{IOLF}(\vec{X},\vec{Y};\vec{w})\right)^2\right] = \vec{w}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)\vec{w}$. Looking at $E\left[\text{IOLF}(\vec{X},\vec{Y};\vec{w})\right]$, $\text{IOLF}(\vec{X},\vec{Y},\vec{w})$ is an unbiased estimator for μ if and only if $\vec{\Delta}^t\vec{w} = 0$, i.e. $\vec{\Delta}$ and $\vec{w}$ are orthogonal. In the following result, two elementary situations where this property holds are provided.

Proposition 4.5 *[35] Let $\vec{X}$ be a random vector with all components having the same mean μ. Let $\vec{Y}$ be a random vector and let $\vec{w} \in \mathbb{R}^n$ be a generalized weighting vector. Then,*

1. *If $\vec{Y}$ and $\vec{X}$ are independent and $\vec{Y}$ is exchangeable, then $IOLF(\vec{X},\vec{Y};\vec{w})$ is an unbiased estimator for μ,*
2. *If the components of $\vec{X}$ are symmetric, independent and have the same distribution, $\vec{X} =_{a.s.} \vec{Y}$ and $\vec{w}$ satisfies $w_k = w_{n-k+1}$ for any $k \in [n-1]$, then $IOLF(\vec{X},\vec{Y};\vec{w})$ is an unbiased estimator for μ.*

Proof: The first statement is a direct consequence of 2 in Proposition 4.3, since in this case $\vec{\Delta} = \vec{0}$.

For the second one, since $\vec{X} =_{a.s.} \vec{Y}$ and the components of $\vec{X}$ are independent and have the same distribution, $\text{IOLF}\left(\vec{X},\vec{Y}\right)$ is a linear combination of the order statistics of $\vec{X}$. Moreover, since the distribution is symmetric, it holds that $\Delta_k = -\Delta_{n-k+1}$ for any $k \in [n]$. Thus, since $w_k = w_{n-k+1}$ for any $k \in [n]$, $\vec{w}^t\vec{\Delta} = 0$. ∎

In general, it is preferred to sacrifice the unbiased property of the estimator to reduce the Mean Squared Error. Since negative weights are allowed, a closed expression of the optimal weights can be found by using Lagrange multipliers.

Theorem 4.6 *[35] Let $\vec{X}$ be a random vector with all components having the same mean μ. Let $\vec{Y}$ be a random vector. Then, the generalized weighting vector $\vec{w} \in \mathbb{R}^n$ which minimizes $E\left[\left(\mu - IOLF(\vec{X},\vec{Y};\vec{w})\right)^2\right]$ is:*

$$\vec{w} = \frac{\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}. \tag{4.4}$$

Proof: Express $E\left[\left(\mu - \text{IOLF}(\vec{X},\vec{Y};\vec{w})\right)^2\right] = \vec{w}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}\right)\vec{w}$ as $\vec{w}^t\Sigma\vec{w} + \left(\vec{w}^t\vec{\Delta}\right)^2$. Then, consider the following optimization problem:

$$\text{Minimize } \vec{w}^t\Sigma\vec{w} + \left(\vec{w}^t\vec{\Delta}\right)^2,$$
$$\text{Subject to } \vec{1}^t\vec{w} = 1.$$

Noticing that $\left(\vec{w}^t\vec{\Delta}\right)^2 = \vec{w}^t\vec{\Delta}\vec{\Delta}^t\vec{w}$ and using Lagrange multipliers [277], the expression is the following

$$\vec{w}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)\vec{w} - \lambda\left(\vec{1}^t\vec{w} - 1\right).$$

Deriving by $\vec{w}$ and making it equal to 0,

$$2\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)\vec{w} - \lambda\vec{1} = 0,\ \vec{w} = \frac{\lambda}{2}\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}.$$

Then, substituting in the restriction,

$$\frac{\lambda}{2}\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1} = 1, \qquad \frac{\lambda}{2} = \frac{1}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}},$$
$$\vec{w} = \frac{\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}.$$

■

The matrix $\Sigma + \vec{\Delta}^t\vec{\Delta}$ is equivalent to a sort of covariance matrix with respect to the real value of the components of the ordered random vector $\pi_{\vec{Y}}(\vec{X})$. Then, negative optimal weights appear if some sums of the rows of $\left(\Sigma^{-1} + \vec{\Delta}^t\vec{\Delta}\right)^{-1}$ are negative. Since the diagonal elements must be positive (since the matrix is positive semi-definite), negative weights imply negative non-diagonal elements of the latter matrix. In the case of covariance matrices, negative elements of the inverse matrix are related to positive conditional correlations and its magnitude to the strength of the dependence [224, 282]. A similar interpretation can be made in this case, negative weights are associated with large positive conditional correlations between the variables of the ordered random vector $\pi_{\vec{Y}}(\vec{X})$.

Notice that if the sample covariance matrix and the sample mean are used to estimate Σ and $\vec{\Delta}$, the here-presented problem is equivalent to Equation 4.3, but without the requirement of positive weights. In addition, this result gives a closed expression of the optimal weights, which is useful both for proving properties and for facilitating calculations. For instance, the Mean Squared Error of the optimal estimator can be computed.

Corollary 4.7 *[35] Let $\vec{X}$ be a random vector with all components having the same mean μ. Let $\vec{Y}$ be a random vector. Then, for all possible values of $\vec{w} \in \mathbb{R}^n$ verifying that $\sum_{i=1}^{n} w_i = 1$, the minimum value of $E\left[\left(\mu - IOLF(\vec{X},\vec{Y};\vec{w})\right)^2\right]$ is*

$$E\left[\left(\mu - IOLF(\vec{X},\vec{Y};\vec{w})\right)^2\right] = \frac{1}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}.$$

Proof: The expectation is straightforward to compute using Theorem 4.6,

$$E\left[\left(\mu - \mathrm{IOLF}(\vec{X},\vec{Y};\vec{w})\right)^2\right] = \vec{w}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}\right)\vec{w} =$$
$$= \frac{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}\left(\Sigma + \vec{\Delta}\vec{\Delta}\right)\frac{\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}} =$$
$$= \frac{1}{\vec{1}^t\left(\Sigma + \vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}.$$

■

Finally, some cases in which the weights computed in Theorem 4.6 lead to an unbiased estimator are provided.

Proposition 4.8 *[35] Let $\vec{X}$ be a random vector with all components having the same mean μ. Let $\vec{Y}$ be a random vector. If one of the following conditions are fulfilled:*

1. $\vec{\Delta} = \vec{0}$,

2. $\vec{X} = \vec{Y}$ *and the components of* $\vec{X}$ *are independent and symmetric.*

Then, the generalized weighting vector with the expression given in Theorem 4.6 makes $IOLF(\vec{X},\vec{Y};\vec{w})$ *an unbiased estimator of* μ.

Proof: Recall that $\vec{w}^t\vec{\Delta} = 0$ is a sufficient condition for the IOLF operator to be an unbiased estimator of μ. If $\vec{\Delta} = \vec{0}$, then $\vec{w}^t\vec{\Delta} = 0$ regardless of the expression of $\vec{w}$.

For the other case, without loss of generality, suppose that $\mu = 0$. The components of $\pi_{\vec{Y}}(\vec{X})$ coincide with the order statistics of $\vec{X}$. Then, $(X_{(i)},X_{(j)}) =_{st} -(X_{(n+1-i)},X_{(n+1-j)})$ and, as a consequence, Σ is persymmetric [142].

Thus, since the inverse of a persymmetric matrix is persymmetric [142], Σ^{-1} is also persymmetric. Then, the weights that minimize the variance, which are $\vec{w} = \frac{\Sigma^{-1}\vec{1}}{\vec{1}\Sigma^{-1}\vec{1}}$ holds that $w_i = w_{n+1-i}$ for any $i \in [n]$. Also, it holds that $\Delta_i = -\Delta_{n+1-i}$ for any $i \in [n]$, Then, $\vec{w}^t\Delta = 0$. The result follows by noticing that if the weights minimize the variance and $\vec{w}^t\Delta = 0$, they also minimize the Mean Squared Error. ■

4.2 L-statistics in mean estimation of symmetric random variables

This section is devoted to providing methods that use extended pre-aggregation functions for mean estimation of symmetric distributions. Recall that aggregations of random variables induced by OWA operators (Definition 2.16) are a particular type of linear combination of order statistics, also known as L-statistics [164].

Consider a quantity of interest $\mu \in \mathbb{R}$ and assume that several measures of this quantity, perturbed by symmetric noise, are given. The result is a sequence of random variables $X_1,\ldots,X_n$ in which all variables are independent, with the same symmetric distribution and mean $\mu \in \mathbb{R}$.

Suppose that one wants to estimate μ using the order statistics of $X_1,\ldots,X_n$. Of course, this approach is a particular case of the one given in Section 4.1 for the IOLF operator. Then, the optimal L-statistic for mean estimation can be found by optimizing the weights as in Theorem 4.6.

Corollary 4.9 *[37] Let $X_1,\ldots,X_n$ be independent and identically distributed random variables with expectation μ. Then, the generalized weighting vector $\vec{w}\in\mathbb{R}^n$ which minimizes $E\left[\left(\mu-\sum_{i=1}^n w_i X_{(i)}\right)^2\right]$ is*

$$\vec{w}=\frac{\left(\Sigma+\vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}{\vec{1}^t\left(\Sigma+\vec{\Delta}\vec{\Delta}^t\right)^{-1}\vec{1}}, \tag{4.5}$$

where Σ denotes the covariance matrix of the order statistics of $X_1,\ldots,X_n$ and $\vec{\Delta}\in\mathbb{R}^n$ is the vector defined as $\Delta_i=E[X_{(i)}]-\mu$ for any $i\in[n]$.

Proof: Apply Theorem 4.6 considering $\vec{Y}=\vec{X}$. ∎

Notice that the resulting operator may have negative weights. Therefore, it is not an aggregation of random variables but a pre-aggregation of random variables, just as the IOLF operator was. In addition, if the noise is multiplied by a scalar quantity, both Σ and $\vec{\Delta}\vec{\Delta}^t$ are multiplied by the square of the quantity, so the optimal weights remain the same.

The optimal weights basically depend on the noise distribution and the sample size, since they characterize the order statistics of the sample. Since the objective is to construct an extended pre-aggregation, it is necessary to explore the relation between weights for the same distribution but different dimension. However, comparing two vectors of different lengths is not straightforward. Inspired by the EVR-OWA theory to derive weights [140], the following function can be defined.

Definition 4.10 *[37] Let $\vec{w}\in\mathbb{R}^n$ be a generalized weighting vector. Then, its cumulative weight function $W:\left\{0,\frac{1}{n},\ldots,\frac{n-1}{n},1\right\}\to\mathbb{R}$ is defined as $W\left(\frac{k}{n}\right)=\sum_{i=1}^k w_i$ for any $k\in[n]$ and $W(0)=0$.*

In the following, the optimal cumulative weight function for a fixed distribution and a sample size n will be denoted as $W^{(n)}$. The graphical representation of the cumulative weights for some classical distributions suggests a sort of convergence. For instance, when working with the hyperbolic secant or the logistic distribution and sample sizes close to 20, the behavior of the cumulative weights seems to be distributed in a common line (see Figure 4.1).

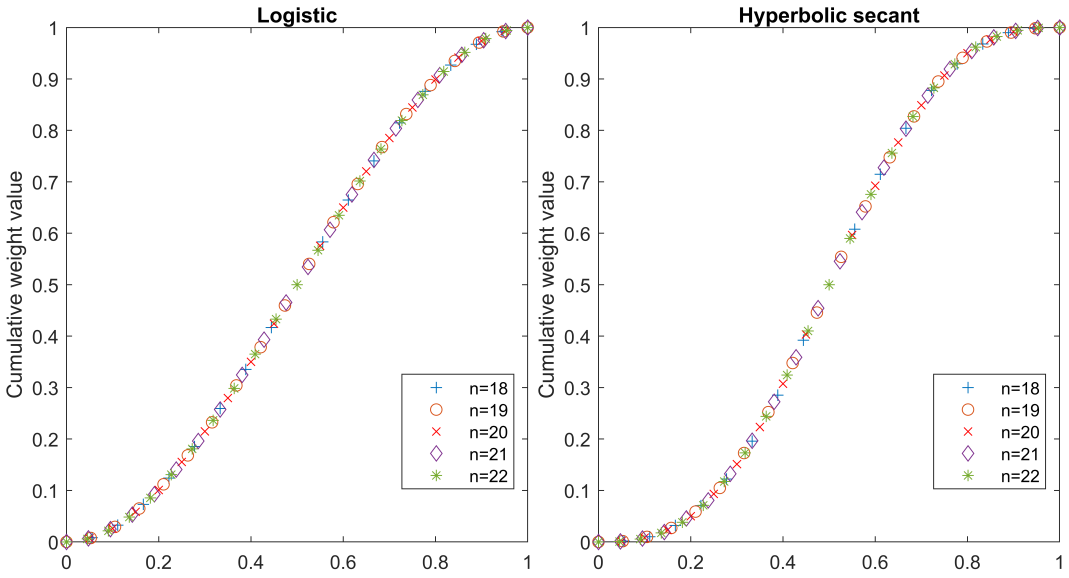

Figure 4.1: Cumulative optimal weights of the logistic (left) and hyperbolic secant (right) distributions for $n \in \{18, 19, 20, 21, 22\}$ [36].

4.2.1 Convergence of optimal cumulative weights

As illustrated in Figure 4.1, the cumulative weights for different sample sizes seem to fit to a common line. This section is devoted to showing that, under certain conditions for a symmetric distribution, the optimal cumulative weights actually converge to a function defined in the interval $[0,1]$. This convergence is based on the convergence of order statistics.

Recall that, roughly speaking, the inverse of the covariance matrix of the order statistics can be approximated as $\Sigma^{-1} \sim (n+1)(n+2)DQD$, Q being the matrix given in Equation 2.1 and D the diagonal matrix satisfying that $D_{i,i} = f\left(F^{-1}(\frac{i}{n+1})\right)$ for any $i \in [n]$. Although this formula is interesting, a formal result regarding this convergence is needed. In particular, if a uniform fast convergence of

$$n\mathrm{Cov}\left(X_{(nq)}, X_{(np)}\right) f\left(F^{-1}(p)\right) f\left(F^{-1}(q)\right),$$

when $p, q \in (0,1)$ is assumed, the elements of Σ^{-1} converge uniformly, and, moreover, it is possible to obtain a dominated sequence if some elements are not considered. The next proposition establishes the convergence of the inverse of the covariance matrix of the order statistics.

Proposition 4.11 *[37] Let $X_1,\ldots,X_n$ be a sequence of continuous, independent and identically distributed random variables with density function $f(x)$ and cumulative distribution $F(x)$ such that $f(x)$ is bounded and strictly positive on $F^{-1}((0,1))$. Suppose that*

$$\lim_{n\to\infty} k(n)(n+1)^2\left((n+2)f\left(F^{-1}(p)\right)f\left(F^{-1}(q)\right)\Sigma_{(n+1)p,(n+1)q}-p(1-q)\right)=0,$$

uniformly for $p,q\in(0,1)$ such that $p\leq q$ for a sequence $k(n)$ such that

$$\lim_{n\to\infty} k(n)\left(\min_{i\in[n]} f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)\right)^2=\infty.$$

Then, for any $\varepsilon>0$ there exists $m\in\mathbb{N}$ such that for any $m\leq n\in\mathbb{N}$,

$$\left|\frac{1}{(n+1)(n+2)}\left(\Sigma^{-1}\right)_{i,j}-(DQD)_{i,j}\right|<\frac{\varepsilon}{k(n)\left(\min_{i\in[n]} f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)\right)^2},$$

for any $i,j\in[n]$, being Q and D the matrices defined previously.

Proof: Consider any $\varepsilon>0$. For a fixed n, the possible values of p and q are, respectively, $p=\frac{i}{n+1}$ and $q=\frac{j}{n+1}$ with $i,j\in[n]$. Using the uniform limit of the hypothesis, for any $\varepsilon_0>0$, there exists $m\in\mathbb{N}$ such that for any $m\leq n\in\mathbb{N}$ the quantity:

$$k(n)(n+1)^2\left|(n+2)f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)f\left(F^{-1}\left(\frac{j}{n+1}\right)\right)\Sigma_{i,j}-\frac{1}{n+1}\left(Q^{-1}\right)_{i,j}\right|,$$

is strictly smaller than ε_0, where $1\leq i\leq j\leq n$ and has been used that $(Q^{-1})_{i,j}=\frac{i(n+1-j)}{n+1}$ for any $i,j\in[n]$. Then, since $f(x)$ is strictly positive on $F^{-1}((0,1))$ the latter expression is equivalent to:

$$\left|(n+2)(n+1)\Sigma_{i,j}-\left((DQD)^{-1}\right)_{i,j}\right|<\frac{\varepsilon_0}{f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)f\left(F^{-1}\left(\frac{j}{n+1}\right)\right)(n+1)k(n)}.$$

Therefore, it is possible to write $(n+2)(n+1)\Sigma=(DQD)^{-1}-\varepsilon_1 X$, where $\varepsilon_1=\frac{\varepsilon_0}{J(n)^2(n+1)k(n)}$, $J(n)$ denotes $\min_{i\in[n]} f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)$, $J(n)>0$ by hypothesis and $\left|X_{i,j}\right|<1$ for any $i,j\in[n]$. Using Lemma 2.2,

$$\frac{1}{(n+2)(n+1)}\Sigma^{-1}=\left((DQD)^{-1}-\varepsilon_1 X\right)^{-1}=$$

$$=\sum_{k=0}^{\infty}(DQD\varepsilon_1 X)^k DQD=DQD+\sum_{k=1}^{\infty}(DQD\varepsilon_1 X)^k DQD.$$

Considering the infinite norm for matrices and denoting as M the supremum of $f(x)$ over $F^{-1}((0,1))$, which exists since f is bounded over $F^{-1}((0,1))$, it is obtained the following inequality for the second term in the previous sum:

$$\left|\left|\sum_{k=1}^{\infty}(DQD\varepsilon_1 X)^k DQD\right|\right|_{\infty} \leq \sum_{k=1}^{\infty}||D||_{\infty}^{2(k+1)}||Q||_{\infty}^{k+1}||X||_{\infty}^{k}\varepsilon_1^k \leq$$
$$\leq 4M^2\sum_{k=1}^{\infty}\left(4nM^2\varepsilon_1\right)^k.$$

Now, set $\varepsilon_0 \leq \frac{\min\{T,1\}\min\{\varepsilon,1\}}{32(\max\{M,1\})^4}$, with $T = \inf_{n\in\mathbb{N}} k(n)J(n)^2$, which exists since $lim_{n\to\infty}k(n)J(n)^2 = \infty$. Therefore,

$$\varepsilon_1 \leq \frac{\min\{T,1\}\min\{\varepsilon,1\}}{32(n+1)k(n)J(n)^2\left(\max\{M,1\}\right)^4},$$

$$\left|\left|\sum_{k=1}^{\infty}(DQD\varepsilon_1 X)^k(DQD)\right|\right|_{\infty} < 4M^2\sum_{k=1}^{\infty}\left(\frac{\min\{T,1\}\min\{\varepsilon,1\}}{8k(n)J(n)^2\left(\max\{M,1\}\right)^2}\right)^k \leq$$

$$\leq 4M^2\frac{\frac{\min\{T,1\}\min\{\varepsilon,1\}}{8k(n)J(n)^2(\max\{M,1\})^2}}{1-\frac{\min\{T,1\}\min\{\varepsilon,1\}}{8k(n)J(n)^2(\max\{M,1\})^2}} \leq \frac{\frac{\varepsilon}{2k(n)J(n)^2}}{1-\frac{1}{8}} = \frac{4}{7k(n)J(n)^2}\varepsilon < \frac{\varepsilon}{k(n)J(n)^2}.$$

The absolute value of any element of a matrix is less than or equal to the infinite norm of the matrix, thus for any $\varepsilon > 0$ there exists $m \in \mathbb{N}$ such that for any $m \leq n \in \mathbb{N}$:

$$\left|\frac{1}{(n+1)(n+2)}\left(\Sigma^{-1}\right)_{i,j} - (DQD)_{i,j}\right| < \frac{\varepsilon}{k(n)J(n)^2},$$

for any $i, j \in [n]$. ∎

In Proposition 4.11, some strong assumptions have been made. The first one is the convergence of the presented sequence. According to Theorem 2.87, the difference must converge to 0, but including the factor $k(n)(n+1)^2$ implies a requirement for faster convergence. In addition, a fast convergence with respect to the vanishing of the density function at the extremes of the support is required. The second one is the uniform convergence rate. Using Theorem 2.87, the uniform convergence on any closed interval contained in $(0,1)$ is guaranteed, but not on $(0,1)$ itself.

The following theorem states the convergence of the cumulative weights, denoted as $\left(W^{(n)}, n \in \mathbb{N}\right)$, when the distribution is fixed and some requirements are fulfilled. More precisely, fast convergence and uniform convergence of the order statistics are needed, as well as some properties of the density function.

Theorem 4.12 *[37] Let $X_1, \ldots, X_n$ be a sequence of continuous, independent and identically distributed random variables with support S, density function f and cumulative distribution F such that f is bounded, continuous, with second derivative on S and strictly positive on $F^{-1}((0,1))$. Suppose that:*

- *There exists a sequence $k(n)$ such that*

$$\lim_{n\to\infty} \frac{k(n)\left(\min_{i\in[n]} f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)\right)}{n^3} = \infty,$$

  satisfying that for any $p, q \in (0,1), p \leq q$:

$$\lim_{n\to\infty} k(n)(n+1)^2\left((n+2)f\left(F^{-1}(p)\right) f\left(F^{-1}(q)\right)\Sigma_{(n+1)p,(n+1)q} - \right.$$
$$\left. - p(1-q)\right) = 0,$$

  uniformly,

- *The integral*

$$\int_0^1 f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2} f\left(F^{-1}(x)\right)\right) dx,$$

  is finite,

- *The limit*

$$\lim_{x\to \inf S} \frac{f(x)}{F(x)}\left(2f(x) - f\left(F^{-1}\left(2F(x)\right)\right)\right),$$

  is not oscillatory.

Then, the sequence of optimal cumulative weights $\left(W^{(n)}, n \in \mathbb{N}\right)$ satisfy $W^{(n)}(0) = 0$, $W^{(n)}(1) = 1$ for any $n \in \mathbb{N}$ and for any $q \in (0,1) \cap \mathbb{Q}$ with irreducible fraction $\frac{a}{b}$:

$$\lim_{n\to\infty} W^{(nb)}(q) = \frac{L + \int_0^q f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2} f\left(F^{-1}(x)\right)\right) dx}{2L + \int_0^1 f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2} f\left(F^{-1}(x)\right)\right) dx},$$

if $\lim_{x\to \inf S} \frac{f(x)}{F(x)} \left(2f(x) - f\left(F^{-1}(2F(x))\right)\right) = L$ *is finite and* $\lim_{n\to\infty} W^{(nb)}(q) = 0.5$ *otherwise.*

Proof: If X is symmetric, similarly as in Proposition 4.8, then $\Delta_i = -\Delta_{n+1-i}$ for any $i \in [n]$, which implies $\vec{w}^t \Delta = 0$. Therefore, the expression of the optimal weights given in Corollary 4.9 can be reduced to $\vec{w} = \frac{\Sigma^{-1}\vec{1}}{\vec{1}^t \Sigma^{-1}\vec{1}}$.

Notice that it is possible to express

$$W^{(n)}\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\Sigma^{-1}\right)_{i,j}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\Sigma^{-1}\right)_{i,j}},$$

for any $k \in [n]$.

For clarity, denote as $\Sigma(n)^{-1}$ the inverse of the covariance matrix of the order statistics of dimension n. Then, consider the sequence $\left(W^{(nb)}(q), n \in \mathbb{N}\right)$ defined as:

$$W^{(nb)}(q) = \frac{\sum_{i=1}^{na}\sum_{j=1}^{nb}\left(\Sigma(nb)^{-1}\right)_{i,j}}{\sum_{i=1}^{nb}\sum_{j=1}^{nb}\left(\Sigma(nb)^{-1}\right)_{i,j}},$$

where $\frac{a}{b}$ is the irreducible fraction of q. Divide by $(nb+1)(nb+2)$ the numerator and the denominator in the latter expression.

Firstly, express the limit of the new numerator as:

$$\lim_{n\to\infty}\left(\sum_{i=1}^{na}\sum_{j=1}^{nb}(DQD)_{i,j} + \left(\sum_{i=1}^{na}\sum_{j=1}^{nb}\frac{\left(\Sigma(nb)^{-1}\right)_{i,j}}{(nb+1)(nb+2)} - (DQD)_{i,j}\right)\right).$$

Secondly, applying Proposition 4.11, there exists $m \in \mathbb{N}$ such that for any $m \leq n \in \mathbb{N}$ the absolute value of the second term can be bounded as follows:

$$\left|\sum_{i=1}^{na}\sum_{j=1}^{nb}\frac{\left(\Sigma(nb)^{-1}\right)_{i,j}}{(nb+1)(nb+2)} - (DQD)_{i,j}\right| \leq \sum_{i=1}^{na}\sum_{j=1}^{nb}\frac{\varepsilon}{k(n)J(n)^2} = \frac{ab\varepsilon n^2}{k(n)J(n)^2},$$

where $J(n) = \left(\min_{i\in[n]} f\left(F^{-1}\left(\frac{i}{n+1}\right)\right)\right)$.

Suppose that this term is negligible in the limit compared to the other terms (this will be proved in (*)). For simplicity, denote $f(F^{-1}(x))$ as ϕ in the next two

equations. Then,

$$\lim_{n\to\infty}\sum_{i=1}^{na}\sum_{j=1}^{nb}(DQD)_{i,j}=\lim_{n\to\infty}\phi\left(\frac{1}{n+1}\right)\left(2\phi\left(\frac{1}{n+1}\right)-\phi\left(\frac{2}{n+1}\right)\right)+$$
$$+\sum_{i=2}^{na}\phi\left(\frac{i}{n+1}\right)\left(2\phi\left(\frac{i}{n+1}\right)-\phi\left(\frac{2}{n+1}\right)-\phi\left(\frac{i-1}{n+1}\right)\right).$$

The limit of the latter line can be expressed in terms of an integral, that is convergent by hypothesis, and the second derivative of f, which is finite:

$$\lim_{n\to\infty}\sum_{i=2}^{na}\phi\left(\frac{i}{n+1}\right)\left(2\phi\left(\frac{i}{n+1}\right)-\phi\left(\frac{i+1}{n+1}\right)-\phi\left(\frac{i-1}{n+1}\right)\right)=$$
$$=\lim_{n\to\infty}\frac{1}{n}\sum_{i=2}^{na}\phi\left(\frac{i}{n+1}\right)\left(\frac{2\phi\left(\frac{i}{n+1}\right)-\phi\left(\frac{i+1}{n+1}\right)-\phi\left(\frac{i-1}{n+1}\right)}{1/n^2}\right)\frac{1}{n}=$$
$$=\lim_{n\to\infty}\frac{1}{n}\int_0^q f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx.$$

Applying the same process to the denominator, the following is obtained:

$$\lim_{n\to\infty}W^{(nb)}(q)=\lim_{n\to\infty}\frac{f\left(F^{-1}\left(\frac{1}{n+1}\right)\right)\left(2f\left(F^{-1}\left(\frac{1}{n+1}\right)\right)-f\left(F^{-1}\left(\frac{2}{n+1}\right)\right)\right)+}{2f\left(F^{-1}\left(\frac{1}{n+1}\right)\right)\left(2f\left(F^{-1}\left(\frac{1}{n+1}\right)\right)-f\left(F^{-1}\left(\frac{2}{n+1}\right)\right)\right)+}$$
$$\frac{+\frac{1}{n}\int_0^q f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx}{+\frac{1}{n}\int_0^1 f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx},$$

where it has been used that $f\left(F^{-1}\left(\frac{1}{n+1}\right)\right)=f\left(F^{-1}\left(\frac{n}{n+1}\right)\right)$ and $f\left(F^{-1}\left(\frac{2}{n+1}\right)\right)=f\left(F^{-1}\left(\frac{n-1}{n+1}\right)\right)$, as a consequence of the symmetry of the distribution. Thus, if

$$\lim_{n\to\infty}nf\left(F^{-1}\left(\frac{1}{n}\right)\right)\left(2f\left(F^{-1}\left(\frac{1}{n}\right)\right)-f\left(F^{-1}\left(\frac{2}{n}\right)\right)\right)=$$
$$=\lim_{x\to\inf S}\frac{f(x)}{F(x)}\left(2f(x)-f\left(F^{-1}(2F(x))\right)\right)=L,$$

the first case of the result is obtained. If the limit diverges, the integral terms are negligible and one has 0.5 as the limit. Notice that it has been assumed that the latter limit cannot be oscillatory.

(*) In order to end the proof only remains to prove that the term bounded by $\frac{ab\varepsilon n^2}{k(n)J(n)^2}$ is negligible with respect to the other terms in the denominator. Notice that the integral can be expressed, integrating by parts:

$$\int_0^1 f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx = \lim_{q\to 1} f\left(F^{-1}(q)\right)\left(\frac{d}{dx}f\left(F^{-1}(q)\right)\right)$$
$$-\lim_{q\to 0} f\left(F^{-1}(q)\right)\left(\frac{d}{dx}f\left(F^{-1}(q)\right)\right) - \int_0^1 \left(\frac{d}{dx}f\left(F^{-1}(x)\right)\right)^2 dx.$$

Applying the symmetry of the distribution, the first and the second terms (with sign) equal the same value. If the latter expression is different from 0, then the term $\frac{ab\varepsilon n^2}{k(n)J(n)^2}$ is negligible with respect to the integral divided by n.

Otherwise, if $f'(x) = 0$ then $f(x)$ is constant and strictly positive. But then $L \neq 0$ (moreover, the associated limit diverges), thus the term $\frac{ab\varepsilon n^2}{k(n)J(n)^2}$ is also negligible. If $f'(x) \neq 0$, then the latter expression is 0 if and only if $\lim_{x\to \inf S} f(x) \neq 0$ and in this case also the limit associated with L diverges. ∎

The previous result only makes sense for the rational numbers on $[0,1]$. However, it can be extended to real numbers over the unit interval straightforwardly by considering its continuous extension.

Theorem 4.12 allows one to expect a smaller difference between the fitting function and the exact solution when the sample size increases. This is very important since, in general, it is not easy to derive the explicit expression of the latter limit because the integrals can be hard to compute. In the next two examples, the limit is computed for the simple cases of the hyperbolic secant and standard logistic distributions.

Example 4.13 *[37] Consider a random variable defined as $Y = \mu + \lambda X$, $\mu, \lambda \in \mathbb{R}$ with X having a hyperbolic secant distribution. The density, distribution and quantile functions of X are, respectively, $f(x) = \frac{1}{2}sech\left(\frac{\pi}{2}x\right)$, $F(x) = \frac{2}{\pi}arctan\left(e^{\frac{\pi}{2}x}\right)$ and $F^{-1}(x) = \frac{2}{\pi}\ln\left(tan\left(\frac{\pi}{2}x\right)\right)$. The limit L equals 0,*

$$\lim_{x\to-\infty} \frac{\pi sech\left(\frac{\pi}{2}x\right)}{4arctan\left(e^{\frac{\pi}{2}x}\right)}\left(sech\left(\frac{\pi}{2}x\right) - \frac{tan\left(2arctan\left(e^{\frac{\pi}{2}x}\right)\right)}{tan^2\left(2arctan\left(e^{\frac{\pi}{2}x}\right)\right)+1}\right) = 0,$$

since the first fraction converges and the second term tends to 0. Then,

$$\frac{d^2}{dx^2}\frac{1}{2}sech\left(\ln\left(tan\left(\frac{\pi}{2}x\right)\right)\right) = \frac{d}{dx}\frac{\pi}{2}\cos(\pi x) = -\frac{\pi^2}{2}\sin(\pi x),$$

and therefore the integral term in Theorem 4.12 is the following:

$$\int_0^x -\frac{\pi^4}{4}sech\left(\ln\left(tan\left(\frac{\pi}{2}t\right)\right)\right)\sin(\pi t)dt = \frac{1}{8\pi}\left(\sin(2\pi x) - 2\pi x\right).$$

It is concluded that the limit function is $g(x) = x - \frac{1}{2\pi}\sin(2\pi x)$.

Example 4.14 *Consider a random variable defined as* $Y = \mu + \lambda X$, $\mu, \lambda \in \mathbb{R}$ *with* X *having a standard logistic distribution. The density, distribution and quantile functions of* X *are, respectively,* $f(x) = \frac{1}{4}sech^2\left(\frac{x}{2}\right)$, $F(x) = \frac{1}{2}\left(1 + tanh\left(\frac{x}{2}\right)\right)$ *and* $F^{-1}(x) = -2tanh^{-1}(1-2x)$. *First, compute the limit L.*

$$\lim_{x\to-\infty}\frac{sech^2\left(\frac{x}{2}\right)}{2\left(1+tanh\left(\frac{x}{2}\right)\right)}\left(\frac{1}{2}sech^2\left(\frac{x}{2}\right) - \frac{1}{2}sech^2\left(-tanh^{-1}\left(1-2tanh\left(\frac{x}{2}\right)\right)\right)\right).$$

Noticing that the limit of the first fraction is 1 and the second term tends to 0, the latter limit equals 0. Secondly, the second derivative of $f\left(F^{-1}(x)\right)$ *should be computed*

$$\frac{d^2}{dx^2}\frac{1}{4}sech^2\left(-tanh^{-1}(1-2x)\right) = \frac{d}{dx}1 - 2x = -2.$$

Finally,

$$\int_0^x -\frac{1}{2}sech^2\left(-tanh^{-1}(1-2x)\right)dt = 2\left(x^3/3 - x^2/2\right).$$

Thus, the limit function is $g(x) = 3x^2 - 2x^3$.

Finally, the asymptotic convergence of the estimators built with the limits of the optimal cumulative weights to the real value of the mean is stated. In particular, the obtained estimator converges in L^2 to μ under the same conditions as in Theorem 4.12.

Proposition 4.15 *[37] Under the conditions of Theorem 4.12, it holds:*

$$\sum_{i=1}^{n}\left(W\left(\frac{i}{n}\right)-W\left(\frac{i-1}{n}\right)\right)X_{(i)}\rightarrow_{L^2}\mu.$$

with $W:[0,1]\rightarrow\mathbb{R}$ *defined as*

$$W(t)=\frac{L+\int_0^t f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx}{2L+\int_0^1 f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx}.$$

if $\lim_{x\rightarrow\inf S}\frac{f(x)}{F(x)}\left(2f(x)-f\left(F^{-1}(2F(x))\right)\right)=L$ *is finite and* $W(0)=0,W(1)=1$ *and* $W(t)=0.5$ *for any* $t\in(0,1)$ *otherwise.*

Proof: Since both the function $W(t)$ and the distribution are symmetric, it is clear that the expectation of the estimator is μ for any sample size. It remains to prove that its variance converges to 0.

Suppose that $\lim_{x\rightarrow\inf S}\frac{f(x)}{F(x)}\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)=L$ is finite. Denote as λ the finite quantity in the denominator of $W(t)$. Then, using the same notation as the previous proof, the variance can be decomposed as follows.

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{n}\left(W\left(\frac{i}{n}\right)-W\left(\frac{i-1}{n}\right)\right)X_{(i)}\right)&=\frac{L^2}{\lambda^2}\Sigma(n)_{1,1}+\Sigma(n)_{n,n}+\\&+\frac{L}{\lambda^2}\sum_{i=2}^{n-1}\left(\Sigma(n)_{1,i}+\Sigma(n)_{n,i}+\Sigma(n)_{i,1}+\Sigma(n)_{i,n}\right)\times\\&\times\left(\int_{\frac{j-1}{n}}^{\frac{j}{n}}f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)+\\&+\frac{1}{\lambda^2}\sum_{i,j=1}^{n}\left(\int_{\frac{i-1}{n}}^{\frac{i}{n}}f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)\times\\&\times\left(\int_{\frac{j-1}{n}}^{\frac{j}{n}}f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)\Sigma(n)_{i,j}\end{aligned}$$

The limit of the three summands is 0.

First summand. If $L=0$, the first summand is 0. Suppose that $L\neq 0$. In that case, using the uniform convergence of the covariance matrix of the order statistics

one has,

$$\lim_{n\to\infty}\frac{L^2}{\lambda^2}\Sigma(n)_{1,1}=\frac{L^2}{\lambda^2}\lim_{n\to\infty}\frac{1}{n+2}\frac{\frac{1}{n}\left(1-\frac{1}{n}\right)}{f\left(F^{-1}\left(\frac{1}{n}\right)\right)^2}\leq$$

$$\leq\frac{L^2}{\lambda^2}\lim_{n\to\infty}\frac{1}{n}\frac{\frac{1}{n}\left(1-\frac{1}{n}\right)}{f\left(F^{-1}\left(\frac{1}{n}\right)\right)^2}=\frac{L^2}{\lambda^2}\lim_{x\to\inf S}F(x)\frac{F(x)(1-F(x))}{f(x)^2}\leq$$

$$\leq\frac{L^2}{\lambda}\lim_{x\to\inf S}\frac{F(x)^2}{f(x)^2}=\frac{L^2}{\lambda^2}\lim_{x\to\inf S}\frac{F(x)^2}{f(x)^2}\frac{\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)^2}{\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)^2}=$$

$$=\frac{1}{\lambda^2}\lim_{x\to\inf S}\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)^2$$

The latter limit is 0 if $\lim_{x\to\inf S}f(x)=0$, which should holds since, if not, $\lim_{x\to\inf S}\frac{f(x)}{F(x)}\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)$ is not finite. A similar reasoning can be applied to $\lim_{n\to\infty}\frac{L^2}{\lambda}\Sigma(n)_{n,n}$, thus the limit of the first summand is 0.

Second summand. If $L=0$, the first second is 0. Suppose that $L\neq 0$. Using the inequality $\Sigma(n)_{1,i}\leq\sqrt{\Sigma(n)_{1,1}\Sigma(n)_{i,i}}$, which holds as a consequence of the positive semi-definiteness of $\Sigma(n)$, and the hyphotesis of uniform convergence of the covariances of the order statistics, it is possible to express

$$\lim_{n\to\infty}\frac{L}{\lambda^2}\sum_{i=2}^{n-1}\Sigma(n)_{1,i}\left(\int_{\frac{j-1}{n}}^{\frac{j}{n}}f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)\leq$$

$$\leq\lim_{n\to\infty}\frac{L\sqrt{\Sigma(n)_{1,1}}}{\lambda^2}\sum_{i=2}^{n-1}\sqrt{\Sigma(n)_{i,i}}\left(\int_{\frac{j-1}{n}}^{\frac{j}{n}}f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)<$$

$$<\lim_{n\to\infty}\frac{L\sqrt{\Sigma(n)_{1,1}}}{\lambda^2\sqrt{n+2}}\int_0^1\sqrt{\frac{p(1-p)}{f\left(F^{-1}(p)\right)^2}}f\left(F^{-1}(p)\right)\left(\frac{d^2}{dp^2}f\left(F^{-1}(p)\right)\right)dp=$$

$$=\lim_{n\to\infty}\frac{L\sqrt{\Sigma(n)_{1,1}}}{\lambda^2\sqrt{n}}\int_0^1\sqrt{p(1-p)}\left(\frac{d^2}{dp^2}f\left(F^{-1}(p)\right)\right)dp\leq$$

$$\leq\lim_{n\to\infty}\frac{L\sqrt{\Sigma(n)_{1,1}}}{\lambda^2\sqrt{n}}\int_0^1\left(\frac{d^2}{dp^2}f\left(F^{-1}(p)\right)\right)dp.$$

By symmetry, the value of the integral is $\lim_{p\to 0}2\left(\frac{d}{dp}f\left(F^{-1}(p)\right)\right)$. This limit is finite since, if S is bounded the derivative of f exists in the minimum of S and, if S is unbounded, it is clear that $\lim_{x\to-\infty}f'(x)=0$. In addition, as proven for the first summand, $\lim_{n\to\infty}L^2\Sigma(n)_{1,1}=0$. Therefore, the second summand equals 0.

Third summand. Changing again the covariances of the order statistics by its limits and the sums for integrals, the following expression is obtained.

$$\lim_{n\to\infty}\frac{1}{\lambda^2}\sum_{i,j=1}^{n}\left(\int_{\frac{i-1}{n}}^{\frac{i}{n}} f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)\times$$
$$\times\left(\int_{\frac{j-1}{n}}^{\frac{j}{n}} f\left(F^{-1}(x)\right)\left(\frac{d^2}{dx^2}f\left(F^{-1}(x)\right)\right)dx\right)\Sigma(n)_{i,j}=\lim_{n\to\infty}\frac{1}{\lambda^2(n+2)}\times$$
$$\times\int_0^1\int_0^1 p(1-q)\left(\frac{d^2}{dp^2}f\left(F^{-1}(p)\right)\right)\left(\frac{d^2}{dq^2}f\left(F^{-1}(q)\right)\right)dpdq\leq$$
$$\leq\lim_{n\to\infty}\frac{1}{\lambda^2 n}\left(\int_0^1\left(\frac{d^2}{dp^2}f\left(F^{-1}(p)\right)\right)dp\right)^2.$$

The integral is convergent by the same previous arguments. Therefore, the limit of the whole expression goes to 0.

It remains the case when $\lim_{x\to\inf S}\frac{f(x)}{F(x)}\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)$ diverges. In this case, the estimator is just $\frac{1}{2}\left(X_{(n)}+X_{(1)}\right)$ and its variance is

$$\frac{1}{4}\left(\Sigma(n)_{1,1}+\Sigma(n)_{1,n}+\Sigma(n)_{n,1}+\Sigma(n)_{n,n}\right).$$

Proceeding similarly as in the first summand of the other case,

$$\lim_{n\to\infty}\frac{1}{4}\Sigma(n)_{1,1}\leq\frac{1}{4}\lim_{x\to\inf S}\frac{F(x)^2}{f(x)^2}=$$
$$=\frac{1}{4}\lim_{x\to\inf S}\left[\frac{F(x)^2}{f(x)^2}\frac{1}{\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)^2}\right]\left(2f(x)-f\left(\left(F^{-1}(2F(x))\right)\right)\right)^2.$$

The part between the brackets goes to 0, since its inverse goes to infinite, and the limit of the other part is finite, thus it is concluded that the limit is 0. The happens for the rest of terms. Since the mean of the estimator converges is μ and the variance converges to 0, the result holds. ∎

The assumptions of the latter results are strong. However, as will be seen in Section 7.2, simulations reveal that convergence is achieved for typical distributions (see Figure 7.1). In addition, estimation methods based on the latter results have a good behavior, as it will be seen in Table 7.3.

4.3 Uniform random capacities

In applied problems [51, 145, 147, 167, 325], the necessity of generating random capacities appears. In this direction, many studies and algorithms have been developed in the literature in this regard [43, 44, 46, 51, 95, 96, 148, 167]. In general, the main objective of these algorithms is to obtain a fast random generation of capacities that is uniform over the set of all capacities. However, there are no well established methods to test if a random generator of capacities is generating them uniformly or not.

In this section, uniform random capacities are studied. Firstly, their use in order to define randomly induced versions of the Choquet and Sugeno integrals is explored. Secondly, many properties of uniform capacities and uniform distributions regarding some families of capacities are studied. These properties will be of interest in Section 7.3 in order to test the uniformity of some random generators of capacities.

A random capacity can be seen as a random vector of dimension 2^n such that each of its components is associated with a subset of $[n]$. Denoting the random capacity as μ, its values can be expressed as $(\mu(\emptyset), \mu(\{1\}), \ldots, \mu([n]))$. Any random capacity μ fulfills, with probability 1, that $\mu(\emptyset) = 0$, $\mu([n]) = 1$ and $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$. A random capacity is said to be uniform if its distribution fulfills Definition 2.70, with A being the set of all capacities of a given dimension. In the following, the values $\mu(\emptyset) = 0$ and $\mu([n]) = 1$ will be ignored, since they are constant, thus it will be considered a random vector of dimension $2^n - 2$.

4.3.1 The randomly induced Choquet and Sugeno integrals

Capacities are used as parameters for some aggregation functions such as the Choquet and Sugeno integrals. Therefore, it is natural to consider randomly induced Choquet C_μ and Sugeno S_μ integrals. If one considers the random capacity to be independent of the input random vector, the following result can be proved.

Proposition 4.16 *Let $(\mu_{\vec{X}}, \vec{X} \in L_I^n)$ be a family of random capacities such that $\mu_{\vec{X}} =_{st} \mu_{\vec{Y}}$ for any $\vec{X}, \vec{Y} \in L_I^n$ with $\vec{X}$ and $\mu_{\vec{X}}$ being independent for any $\vec{X} \in L_I^n$.*

Then, the function $A : L_I^n \to L_I$ defined by $A(\vec{X}) = C_{\mu_{\vec{X}}}(\vec{X})$ is an aggregation of random variables.

Proof: Consider Theorem 3.22 with $(\vec{\lambda}_{\vec{X}}, \vec{X} \in L_I^n) = (\mu_{\vec{X}}, \vec{X} \in L_I^n)$.

Firstly, it is necessary to prove that $C_{\mu_{\vec{X}}}(\vec{X})$ is a random variable. Consider the probability space $(\Omega, \mathscr{F}, P)$ and $B \in \mathbb{B}$. The sets $M_\pi = \{\omega \in \Omega \mid X_{\pi(1)}(\omega) \leq \cdots \leq X_{\pi(n)}(\omega)\}$ are measurable for any permutation π [93]. For a particular permutation, the Choquet integral equals the function $\sum_{i=1}^{n} \left(X_{\pi(i)} - X_{\pi(i-1)}\right) \mu_{\vec{X}}(A_{\pi(i)})$ (see Definition 2.27). Since the function is defined in terms of sums and multiplications of measurable functions,

$$\left(\sum_{i=1}^{n} \left(X_{\pi(i)} - X_{\pi(i-1)}\right) \mu_{\vec{X}}(A_{\pi(i)})\right)^{-1} (B) \in \mathscr{F}.$$

Considering all the permutations,

$$\left(C_{\mu_{\vec{X}}}(\vec{X})\right)^{-1} (B) = \bigcup_{\pi} \left(\left(\left(\sum_{i=1}^{n} \left(X_{\pi(i)} - X_{\pi(i-1)}\right) \mu_{\vec{X}}(A_{\pi(i)}) \right)^{-1} (B) \right) \cap M_\pi \right),$$

which is a measurable set since it is the result of finite unions and intersections of measurable sets.

Then, notice that Condition (1) in Theorem 3.22 is fulfilled. Condition (2) holds noticing that the Choquet Integral is internal for any possible value of μ. Condition (3) holds by hypothesis and Condition (4) is fulfilled using the independence between $\vec{X}$ and $\mu_{\vec{X}}$ for any $\vec{X} \in L_I^n$. Then, it is concluded that A is an aggregation of random variables. ∎

Using a similar proof, an equivalent result can be stated for the Sugeno integral.

Proposition 4.17 *Let $(\mu_{\vec{X}}, \vec{X} \in L_I^n)$ be a family of random capacities such that $\mu_{\vec{X}} =_{st} \mu_{\vec{Y}}$ for any $\vec{X}, \vec{Y} \in L_I^n$ with $\vec{X}$ and $\mu_{\vec{X}}$ being independent for any $\vec{X} \in L_I^n$. Then, the function $A : L_I^n \to L_I$ defined by $A(\vec{X}) = S_{\mu_{\vec{X}}}(\vec{X})$ is an aggregation of random variables.*

Proof: The proof is analogous to the one of Proposition 4.16. In this case, the expression of the Sugeno integral for a fixed permutation π is

$$S_\mu(\vec{x}) = \max_{i\in[n]} \left(\min\left(x_{\pi(i)}, \mu_{\vec{X}}(A_{\pi(i)})\right)\right),$$

and again one has that

$$\left(\max_{i\in[n]} \left(\min\left(x_{\pi(i)}, \mu_{\vec{X}}(A_{\pi(i)})\right)\right)\right)^{-1}(B) \in \mathscr{F},$$

since it can be decomposed in terms of maximums and minimums of measurable functions [93]. The rest of the proof is equivalent. ∎

4.3.2 Properties of uniform capacities

The uniform distribution over the set of capacities is not simple to study, since the geometry of that set is difficult to work with. The set of all capacities of dimension n, which is a 2^n-2 dimensional order polytope, will be denoted as P_n. All its vertices are 0-1 capacities. One description of this distribution can be done in terms of the set of linear extensions.

Lemma 4.18 *[27] Let μ be a random capacity of dimension n. Then, its density function is $f_\mu(\vec{x}) = \frac{|L_e|}{(2^n-2)!}$ if $\vec{x} \in P_n$ and $f_\mu(\vec{x}) = 0$ otherwise.*

Proof: Divide the unit hyper-cube of dimension 2^n-2 in the possible $(2^n-2)!$ orderings of the values. Each subset has a measure of $\frac{1}{(2^n-2)!}$. Then, add the measures of all the possible linear extensions to obtain that the measure of P_n is $\frac{|L_e|}{(2^n-2)!}$. ∎

However, as explained in Section 2.2.4, the number of linear extensions does not have a closed algebraic expression. Therefore, the latter expression of the density function is not operational. In the next result, a characterization in terms of the order statistics of the random capacity is given.

Proposition 4.19 *[27] Let μ be a random capacity of dimension n. Then, it has uniform distribution if and only if the following conditions are fulfilled:*

- *All linear extensions appear with the same probability,*
- *The order statistics of μ and the linear extension are independent,*
- *The order statistics of μ have the same distribution as the order statistics of a vector of 2^n-2 independent standard uniform random variables.*

Proof: Firstly, prove that the conditions are necessary. Denote as $\mu|\pi$ the (random) value of the random uniform capacity μ when the linear extension π is fixed.

- Since the distribution is uniform, for any linear extension $\pi \in L_e$, $P(\pi)$ equals the Lebesgue measure associated with the linear extension divided by the Lebesgue measure of P_n. This value, $\frac{1}{|L_e|}$, is the same for each possible permutation.
- Computing the density function of μ given a linear extension $\pi \in L_e$,

$$f_{\mu|\pi}(\vec{x}) = \frac{f_\mu(\vec{x})}{P(\pi)} = \frac{(2^n-2)!}{|L_e|}|L_e| = (2^n-2)!,$$

  for each $\vec{x} \in [0,1]^{2^n-2}$ with $x_{\pi(1)} \leq \cdots \leq x_{\pi(2^n-2)}$ and 0 otherwise.

  Denote as $o(\mu|\pi)$ the order statistics of μ given the linear extension π. Since the order of the elements of $\mu|\pi$ is given by π, then $o(\mu|\pi)$ is a non-random permutation of the values of $\mu|\pi$. Then, the associated density function, $f_{o(\mu|\pi)}$, equals the one of the order statistics of 2^n-2 standard random variables [104]. Therefore,

$$f_{o(\mu|\pi)}(\vec{x}) = (2^n-2)!,$$

  for each $\vec{x} \in [0,1]^{2^n-2}$ with $x_1 \leq \cdots \leq x_{2^n-2}$ and 0 otherwise.

  Therefore, the conditional distribution of the order statistics of μ given a linear extension is always the same, and therefore, the order statistics of μ and the linear extension are independent.

- Since the order statistics of μ and the linear extension are independent, the latter conditional density is the actual density of the order statistics of μ, which is the one of the order statistics of a collection of 2^n-2 independent standard uniform random variables.

It remains to prove that these conditions are sufficient. The density function of μ can be expressed as

$$f_\mu(\vec{x}) = \sum_{\pi \in L_e} P(\pi) f_{\mu|\pi}(\vec{x}).$$

The density function of the order statistics of a collection of $2^n - 2$ independent standard uniform random variables is $(2^n - 2)!$. Using that the linear extension and the order statistics of μ are independent, one has that

$$f_{o(\mu|\pi)}(\vec{x}) = (2^n - 2)!,$$

for each $\vec{x} \in [0,1]^{2^n-2}$ with $x_1 \leq \cdots \leq x_{2^n-2}$ and 0 otherwise.

Using again that, given a linear extension, the order statistics are just a deterministic permutation of the values of the fuzzy measure,

$$f_{\mu|\pi}(\vec{x}) = (2^n - 2)!.$$

for each $\vec{x} \in [0,1]^{2^n-2}$ with $x_{\pi(1)} \leq \cdots \leq x_{\pi(2^n-2)}$ and 0 otherwise.

Finally, using that all linear extensions have the same probability, which should be $\frac{1}{|L_e|}$,

$$f_\mu(\vec{x}) = \sum_{\pi \in L_e} P(\pi) f_{\mu|\pi}(\vec{x}) = \frac{(2^n - 2)!}{|L_e|},$$

for each $\vec{x} \in P_n$. ∎

Trivially, the second and third conditions can be combined into one.

Corollary 4.20 *[27] Let μ be a random capacity of dimension n. Then, it has uniform distribution if and only if the following conditions are fulfilled:*

- *All linear extensions appear with the same probability,*
- *Given any linear extension, the order statistics of μ have the same distribution as the order statistics of a vector of $2^n - 2$ independent standard uniform random variables.*

4.3.2.1 Symmetry and inequality properties

Using the characterizations of the latter section, several properties related to the symmetry and inequalities between the measure of subsets can be proved.

Proposition 4.21 *[27] Let μ be a random capacity of dimension n with uniform distribution. Then, the following conditions are fulfilled:*

- *The random vector $(\mu(\{1\}),\ldots,\mu(\{n\}))$ is exchangeable,*
- *$\mu(A) =_{st} \mu(B)$ for each $A,B \subseteq [n]$ with $|A| = |B|$,*
- *$\mu(A) \leq_{st} \mu(B)$ for each $A,B \subseteq [n]$ with $|A| \leq |B|$,*
- *The dual capacity of μ has the same distribution as μ.*

Proof:

- Consider the function $\pi_{i,j} : P_n \to P_n$ that, for fixed $i,j \in [n]$, takes a capacity and returns another capacity following the formula

$$\pi_{i,j}(\mu)(A) = \begin{cases} \mu(A) & \text{if } i \in A, j \in A, \\ \mu(A) & \text{if } i \notin A, j \notin A, \\ \mu(\{j\} \cup (A \backslash \{i\})) & \text{if } i \in A, j \notin A, \\ \mu(\{i\} \cup (A \backslash \{j\})) & \text{if } i \notin A, j \in A. \end{cases}$$

  Since the latter function is just a permutation of two elements of $[n]$ in every subset, it is a Lebesgue-measure preserving bijection on the polytope P_n. Since μ has uniform distribution, which is just the Lebesgue measure scaled by a constant, $\pi_{i,j}(\mu)$ has the same distribution as μ. The random vector $(\pi_{i,j}(\mu)(\{1\}),\ldots,\pi_{i,j}(\mu)(\{n\}))$ is the random vector $(\mu(\{1\}),\ldots,\mu(\{n\}))$ when the components i and j are interchanged. Since $\pi_{i,j}(\mu)$ and μ have the same distribution, the same holds for $(\pi_{i,j}(\mu)(\{1\}),\ldots,\pi_{i,j}(\mu)(\{n\}))$ and $(\mu(\{1\}),\ldots,\mu(\{n\}))$. Since this is true for any $i,j \in [n]$, the random vectors are exchangeable.

- Consider $A, B \subseteq [n]$ with $|A| = |B|$. Since they have the same cardinality, $A \backslash (A \cap B)$ and $B \backslash (A \cap B)$ have the same cardinality, thus they can be represented as $A \backslash (A \cap B) = \{a_1, \ldots, a_k\}$ and $B \backslash (A \cap B) = \{b_1, \ldots, b_k\}$ with $k = |A| - |A \cap B|$. Proceeding as in the latter case, it holds that $\mu(B) = \pi_{a_1,b_1} \circ \cdots \circ \pi_{a_k,b_k}(\mu)(A)$ has the same distribution as $\mu(A)$.

- Consider $A, B \subseteq [n]$ such that $|A| \leq |B|$. Then, there exists $B' \subseteq [n]$ such that $|B'| = |B|$ and $A \subseteq B'$. Since $A \subseteq B'$, it is clear that for any linear extension $\mu(A) \leq \mu(B')$. Therefore, $P(\mu(A) \leq \mu(B')) = 1$, which implies $\mu(A) \leq_{st} \mu(B')$. Using the previous point, $\mu(B') =_{st} \mu(B)$, so it is concluded that $\mu(A) \leq_{st} \mu(B') =_{st} \mu(B)$.

- Consider the function $\phi : P_n \to P_n$ that takes a capacity and returns the dual of the capacity, that is, $\phi(\mu)(A) = 1 - \mu(\bar{A})$ for any $A \subseteq [n]$. This function is a biyection [48]. In addition, since the function is the result of the composition of a permutation of the indices associated with a subset of $[n]$ and its complementary and a biyective linear function, which preserves the Lebesgue measure (see Chapter 11, Section 2.1 in [111]), ϕ preserves the Lebesgue measure. Then ϕ is a Lebesgue-measure preserving bijection on P_n. The property is established in a similar way to the previous cases. ∎

It should be remarked that exchangeability can only be achieved between subsets of the same cardinality when the cardinal is 1 and $n-1$ (see Corollary 4.22). For instance, if the cardinality is 2, the dependence between $\mu(\{1,2\})$ and $\mu(\{1,3\})$ is not the same as the dependence between $\mu(\{1,2\})$ and $\mu(\{3,4\})$, since the intersection of $\{1,2\}$ and $\{1,3\}$ and the intersection of $\{1,2\}$ and $\{3,4\}$ have different cardinality.

As a direct consequence of the last result, the following corollaries arise.

Corollary 4.22 *[27] Let μ be a random capacity of dimension n with uniform distribution. Then, the random vector $(\mu([n]\backslash\{1\}), \ldots, \mu([n]\backslash\{n\}))$ is exchangeable.*

Corollary 4.23 *[27] Let μ be a random capacity of dimension n with uniform distribution. Then, the capacity $E[\mu]$ is symmetric and autodual.*

The expectation of a uniform random capacity coincides with the centroid of the polytope P_n. The last result can also be found in [237], described in these terms.

4.3.2.2 Distribution of the measure of subsets

The next results aim to determine the type of distribution each of the random variables $\mu(A)$ has. Recall the definition of mixture given in Definition 2.65.

Proposition 4.24 *[27] Let μ be a random capacity of dimension n with uniform distribution. Then, the distribution of $\mu(A)$ is a mixture of Beta distributions for any $A \subset [n]$ with $A \neq \emptyset$.*

Proof: Fix $A \subset [n]$ with $A \neq \emptyset$. For each permutation $\pi : [2^n] \to \mathscr{P}([n])$ in the set of linear extensions, $\mu(A)$ behaves as the $\pi^{-1}(A)$-th order statistic in a sample of $2^n - 2$ independent standard uniform random variables. In particular,

$$[\mu(A)|\pi] \sim Beta(\pi^{-1}(A), 2^n - 1 - \pi^{-1}(A)),$$
$$f_{\mu(A)|\pi}(x) = \frac{x^{\pi^{-1}(A)-1}(1-x)^{2^n-2-\pi^{-1}(A)}}{B(\pi^{-1}(A), 2^n - 1 - \pi^{-1}(A)))}, \; x \in [0,1].$$

where $B(\cdot,\cdot)$ denotes the Beta function.

Thus, the total density is of the form

$$f_{\mu(A)}(x) = \frac{1}{|L_e|} \sum_{\pi \in L_e} \frac{x^{\pi^{-1}(A)-1}(1-x)^{2^n-2-\pi^{-1}(A)}}{B(\pi^{-1}(A), 2^n - 1 - \pi^{-1}(A)))}, \; x \in [0,1],$$

which is a mixture of Beta distributions. ∎

In general, since the exact distribution of the measure of the subsets is not known (it is a mixture of Beta distributions but the parameters are not known), it is hard to compute the distribution of associated quantities. However, it can be stated that the Orness has a symmetric distribution.

Corollary 4.25 *[27] Let μ be a random capacity of dimension n with uniform distribution. Then, $O(C_\mu)$ has a symmetric distribution with respect to* 0.5.

Proof: The sum of the Orness of a capacity and its dual is always one. Since μ has the same distribution as its dual random capacity $\hat{\mu}$, one has that $O(C_\mu) =_{st} 1 - O(C_{\hat{\mu}}) =_{st} 1 - O(C_\mu)$. Then, $0.5 - O(C_\mu) =_{st} O(C_\mu) - 0.5$. ∎

This, of course, implies that $E[O(C_\mu)] = 0.5$, a feature that was already known by the fact that $E[\mu]$ is autodual.

4.3.2.3 Distance to the closest 0-1 capacity

More involved results can be proved for the smallest distance from the uniform random capacity to the closest 0-1 capacity. Notice that 0-1 capacities are the (only) vertices of the polytope of capacities, thus being of special interest. In the next lemma, the closest 0-1 capacity to a general capacity is characterized.

Lemma 4.26 *[27] Let μ be a capacity of dimension n. Denote as $P_n^{0,1}$ the set of 0-1 capacities of dimension n, then:*

$$\min_{\hat{\mu} \in P_n^{0,1}} \left(\sum_{A \subseteq [n]} |\mu(A) - \hat{\mu}(A)| \right) = \sum_{A \subseteq [n]} (0.5 - |\mu(A) - 0.5|) .$$

Proof: Consider the 0-1 capacity defined as $\hat{\mu}(A) = 0$ if $\mu(A) \leq 0.5$ and $\hat{\mu}(A) = 1$ if $\mu(A) > 0.5$. It is a capacity since $\hat{\mu}(\emptyset) = 0$, $\hat{\mu}([n]) = 1$ and, by the monotonicity of μ, the case $\hat{\mu}(A) = 0$, $\hat{\mu}(B) = 1$ with $B \subseteq A$ is not possible. This 0-1 capacity is the closest since, by construction, the closest value between 0 and 1 for each subset is being chosen. ∎

The distribution of the latter expression can be computed by using the Irwin-Hall distribution (see Definition 2.71).

Proposition 4.27 *[27] Let μ be a random capacity of dimension n with uniform distribution. Then, the (Manhattan) distance to the closest 0-1 capacity has the distribution of an Irwin-Hall random variable with $n = 2^n - 2$ divided by 2.*

Proof: By Lemma 4.26, given a linear extension π, one has:

$$\sum_{A \subseteq [n]} (0.5 - |[\mu(A)|\pi] - 0.5|) = \sum_{i=1}^{2^n - 2} \left(0.5 - \left|X_{(i)} - 0.5\right|\right),$$

where $X_1,\ldots,X_{2^n-2}$ are independent random variables with standard uniform distribution. Changing the order of the summands,

$$\sum_{A\subseteq[n]}(0.5-|[\mu(A)|\pi]-0.5|)=\sum_{i=1}^{2^n-2}(0.5-|X_i-0.5|)\,.$$

The distribution of $|X_i-0.5|$ is uniform in $[0,0.5]$ as well as the distribution of $0.5-|X_i-0.5|$. That is, for any linear extension, the minimum distance is the sum of 2^n-2 independent uniform random variables on $[0,0.5]$. The result holds by realizing that multiplying by 2 an Irwin-Hall distribution is obtained, see Proposition 2.72. ∎

Finally, it is possible to compute the probability that a 0-1 capacity is the closest 0-1 capacity to a uniform random capacity.

Proposition 4.28 *[27] Let μ be a uniform capacity and μ' a 0-1 capacity. Then, the probability of μ' being the closest 0-1 capacity to μ is*

$$0.5^{2^n-2}\frac{(2^n-2)!}{S!(2^n-2-S)!}\frac{|L'_e|}{|L_e|},$$

where L'_e is the set of linear extensions of μ' and S is the number of non-empty subsets with measure μ' equal to 0.

Proof: If the linear extension of μ is not in L'_e, then the probability that μ' is the closest 0-1 measure is 0. In other case, consider a fixed $\pi\in L'_e$. Applying Proposition 4.19, the conditional probability given π can be computed as:

$$\begin{gathered}P\left(X_{(1)}\leq 0.5,\ldots,X_{(S)}\leq 0.5,X_{(S+1)}>0.5,\ldots,X_{(2^n-2)}>0.5\right)=\\=P\left(X_{(S)}\leq 0.5,X_{(S+1)}>0.5\right)=\\=P\left(X_{(S)}\leq 0.5,X_{(S+1)}\leq 1\right)-P\left(X_{(S)}\leq 0.5,X_{(S+1)}\leq 0.5\right)=\\=P(X_{(S)}\leq 0.5)-P(X_{(S+1)}\leq 0.5),\end{gathered}$$

where $X_{(1)},\ldots,X_{(2^n-2)}$ are the order statistics of a sample of 2^n-2 independent uniform random variables. By using the distribution function of the order statistics

given in Proposition 2.83,

$$P(X_{(S)} \leq 0.5) - P(X_{(S+1)} \leq 0.5) =$$
$$= \sum_{i=S}^{2^n-2} \frac{(2^n-2)!}{i!(2^n-2-i)!} 0.5^i 0.5^{2^n-2-i} - \sum_{i=S+1}^{2^n-2} \frac{(2^n-2)!}{i!(2^n-2-i)!} 0.5^i 0.5^{2^n-2-i} =$$
$$= 0.5^{2^n-2} \frac{(2^n-2)!}{S!(2^n-2-S)!}.$$

The result follows by doing the average between all the possible linear extensions. ∎

4.3.3 More properties for some particular families of capacities

The set of all capacities is hard to deal with, mainly because of the difficulty of working with the linear extensions. In this section, the study is restricted to some families of capacities in which their characteristics allow one to prove more results.

4.3.3.1 Balanced capacities

For balanced capacities, the monotonicity is extended to a stronger condition, since the measure of subsets is ordered with respect to the cardinality. This allows us to compute the exact number of possible linear extensions.

Proposition 4.29 *[27] Let L_e^b be the set of the linear extensions of balanced capacities of dimension n. Then,*

$$|L_e^b| = \prod_{k=1}^{n-1} \left(\frac{n!}{k!(n-k)!} \right)!.$$

Proof: Any possible linear extension of a balanced capacity can be expressed as $n-1$ permutations of the measures of subsets of the same cardinality since $A \subseteq B \implies \mu(A) \leq \mu(B)$. For a cardinality $k \in [n-1]$, there exist $\frac{n!}{k!(n-k)!}$ different subsets of $[n]$ with that cardinality. The number of permutations of a collection of $\frac{n!}{k!(n-k)!}$ elements is $\left(\frac{n!}{k!(n-k)!}\right)!$. The result follows by considering the combination of all possible permutations of each of the subsets with the same cardinality. ∎

Let μ_b be a random capacity with uniform distribution over the set of all balanced capacities. Its density, proceeding as in the general case, has the following expression:

$$f_\mu(\vec{x}) = \frac{(2^n-2)!}{|L_e^b|} = \frac{(2^n-2)!}{\prod_{k=1}^{n-1}\left(\frac{n!}{k!(n-k)!}\right)!},$$

for any $\vec{x} \in P_n^b$, where P_n^b is the polytope of balanced capacities.

Moreover, it is also possible to know all the positions of a subset in the linear extensions.

Lemma 4.30 *[27] Consider $A \subset [n]$ with $|A| = m$. Then, the position of A over all the linear extensions is equally distributed between the positions $m_l + 1$ and m_u, with expressions*

$$m_l = \sum_{k=1}^{m-1} \frac{n!}{k!(n-k)!},\ m_u = \sum_{k=1}^{m} \frac{n!}{k!(n-k)!}.$$

Proof: It will be proved by induction. For $m = 1$, the position of A should be between the first n subsets, that is, between $1_l + 1 = \sum_{k=1}^{0} \frac{n!}{k!(n-k)!} + 1 = 1$ and $1_u = \sum_{k=1}^{1} \frac{n!}{k!(n-k)!} = \frac{n!}{1!(n-1)!} = n$. Suppose that it is true for $m-1$. Then, since the subsets of cardinality m should be between subsets of cardinality $m-1$ and $m+1$ in any linear extension, one has that $m_l = (m-1)_l + \frac{n!}{(m-1)!(n-m+1)!}$ and $m_u = (m-1)_u + \frac{n!}{m!(n-m)!}$, thus the result holds. ∎

Notice that $\sum_{k=1}^{n-1} \frac{n!}{k!(n-k)!} = 2^n - 2$. Using the latter results, the exact distribution of the measure of each subset can be determined.

Proposition 4.31 *[27] Let μ_b be a uniform balanced random capacity of dimension n and $A \subset [n]$ such that $A \neq \emptyset$. Then, the density function of $\mu_b(A)$ has the following expression:*

$$f_{\mu_b(A)}(x) = \frac{|A|!(n-|A|)!}{n!} \sum_{i=\sum_{k=1}^{|A|-1}\frac{n!}{k!(n-k)!}+1}^{\sum_{k=1}^{|A|}\frac{n!}{k!(n-k)!}} \frac{x^{i-1}(1-x)^{2^n-2-i}}{B(i, 2^n-1-i)},$$

where $B(\cdot,\cdot)$ denotes the Beta function.

Proof: Proceeding as in Proposition 4.24, it is obtained that

$$f_{\mu_b(A)}(x) = \frac{1}{|L_e^b|} \sum_{\pi \in L_e^b} \frac{x^{\pi^{-1}(A)-1}(1-x)^{2^n-2-\pi^{-1}(A)}}{B(\pi^{-1}(A), 2^n-1-\pi^{-1}(A)))}.$$

Using Lemma 4.30, the equation simplifies to the following expression,

$$f_{\mu_b(A)}(x) = \frac{|A|!(n-|A|)!}{n!} \sum_{i=\sum_{k=1}^{|A|-1} \frac{n!}{k!(n-k)!}+1}^{\sum_{k=1}^{|A|} \frac{n!}{k!(n-k)!}} \frac{x^{i-1}(1-x)^{2^n-2-i}}{B(i, 2^n-1-i)}.$$

■

The latter density consists of a mixture of $\frac{n!}{|A|!(n-|A|)!}$ Beta distributions, each of them associated with one order statistic of a sample of 2^n-2 uniform random variables in the positions between $\sum_{k=1}^{|A|-1} \frac{n!}{k!(n-k)!}+1$ and $\sum_{k=1}^{|A|} \frac{n!}{k!(n-k)!}$. In addition, the exact distribution of the Orness can be derived for this family.

Proposition 4.32 *[27] Let μ_b be a uniform random balanced capacity of dimension n. Then, $O(C_\mu)$ has the following distribution function:*

$$F_{O(C_\mu)}(x) = 1 - \sum_{i=1}^{m} \frac{(c_i - x)^n}{c_i \prod_{j \neq i}(c_i - c_j)} \mathbb{I}_{x \in [0, c_1]},$$

where $c_i = \sum_{k=i}^{n} a_k$, $a_i = \frac{1}{n-1} \frac{m!(n-m)!}{n!}$ if $i \in \left\{ \sum_{k=1}^{m-1} \frac{n!}{k!(n-k)!} + 1, \ldots, \sum_{k=1}^{m} \frac{n!}{k!(n-k)!} \right\}$, $\mathbb{I}$ denotes the indicator function and m is the largest integer such that $x \leq c_m$.

Proof: Since for all the linear extensions the subsets are ordered with respect to its cardinality and the coefficients in the linear combination of the Orness only depend on the cardinality, one has that $[O(C_\mu)|\pi]$ has the same distribution for each $\pi \in L_e^b$.

For a particular linear extension, therefore for all of them, using the third point of Proposition 4.19 and Lemma 4.30, one has

$$\frac{1}{n-1} \sum_{A \subset [n]} \frac{(n-|A|)!|A|!}{n!} \mu(A) =_{st} a_i X_{(i)},$$

where $X_{(1)},\ldots,X_{(2^n-2)}$ are the order statistics of a sample of independent uniform random variables and the coefficients $a_1,\ldots,a_n$ are defined as:

$$a_i = \frac{1}{n-1}\frac{m!(n-m)!}{n!}, \text{ if } i \in \left\{\sum_{k=1}^{m-1}\frac{n!}{k!(n-k)!}+1,\ldots,\sum_{k=1}^{m}\frac{n!}{k!(n-k)!}\right\}.$$

Then, since all a_i are strictly positive, apply Proposition 2.85 to obtain the result. ∎

A similar result can be stated when computing the non-modularity index of the whole set. In this case, it is necessary to compute the distribution of the following random variable.

$$1-\frac{1}{n}\sum_{i\in A}\mu_b(\{i\})+\mu_b([n]\backslash\{i\}).$$

Notice that the latter expression is a linear combination with negative coefficients plus a constant. The main limitation regarding reproducing the last result is that there are coefficients of associated linear combination of the order statistics that are 0.

Proposition 4.33 *[27] Let μ_b be a uniform random balanced capacity of dimension n. Then, $d_{\mu_b}([n])$ has the following distribution function,*

$$F_{d_{\mu_b}([n])}(x) = \sum_{i=1}^{m}\frac{g_i^{(n_1-n_{i-1}-1)}(c_{n_i},1-x)}{(n_i-n_{i-1}-1)!}\mathbb{I}_{x\in[1-c_1,1]},$$

where $c_i=\sum_{k=i}^{n}a_k$ with $a_i=\frac{1}{n}$ if $i\in\{1,\ldots,n,2^n-2-n+1,\ldots,2^n-2\}$, $\mathbb{I}$ denotes the indicator function, m is the largest integer such that $1-x\leq c_m$ and $g_i^{(s)}(c,x)$ denotes

$$g_i^{(s)}(c,x)=\frac{\partial^s}{\partial c^s}\frac{(c-(1-x))^n}{c\prod_{j\neq i}(c-c_j)}.$$

Proof: Proceeding similarly to Proposition 4.32 but using Proposition 2.86 instead of Proposition 2.85, the distribution function of $\frac{1}{n}\sum_{i\in A}\mu_b(\{i\})+\mu_b([n]\backslash\{i\})=1-d_{\mu_b}(N)$ is

$$F_{1-d_{\mu_b}(N)}(x)=1-\sum_{i=1}^{m}\frac{g_i^{(n_1-n_{i-1}-1)}(c_{n_i},x)}{(n_i-n_{i-1}-1)!}\mathbb{I}_{x\in[0,c_1]},$$

The result holds by realizing that, if X is a continuous random variable, then $F_{1-X}(x) = 1 - F_X(1-x)$. ∎

4.3.3.2 Belief measures

The family of belief measures is easy to deal with, since it has a convenient characterization in terms of the Mobius transform.

Proposition 4.34 *[48, 146] A capacity μ is a belief measure if and only if $M_\mu(A) \geq 0$ for any $A \subseteq [n]$.*

Recall that if a random vector has a uniform distribution over a subset, any linear bijective transformation of such a random vector also has uniform distribution on the image of the subset, (see Chapter 11 in [111]). Therefore, the Mobius transform of a uniform random belief measure has uniform distribution over the set of Mobius transforms with positive values. Its density function can be computed.

Proposition 4.35 *[27] Let μ_{bel} be a uniform random belief measure. Then, the density function of the Mobius transform of μ_{bel} is $f_{M(\mu_{bel})}(\vec{x}) = (2^n-2)!\mathbb{I}_{\sum_{i=1}^n x_i=1,\ \vec{x}>\vec{0}}$.*

Proof: First, notice that the transformation between the usual representation and the Mobius transform is linear, thus a random belief measure has uniform distribution if and only the Mobius transform also has uniform distribution.

The volume of the set of Mobius transforms of belief measures is defined as the set of 2^n-1 positive values (each one associated with one of the non-null subsets of N) such that their sum equals 1.

The volume (as a subspace of $\mathbb{R}^{2^n-2}$) equals the density function of an Irwin-Hall variable with $x = 1$ and $m = 2^n - 1$. That is:

$$f_{S_{2^n-2}}(1) = \frac{1}{(2^n-2)!}\sum_{k=0}^{\lfloor 1 \rfloor}(-1)^k \frac{(2^n-1)!}{k!(2^n-1-k)!}(1-k)^{2^n-1} =$$

$$= \frac{1}{(2^n-2)!}\left(\frac{(2^n-1)!}{0!(2^n-1)!}(1-0)^{2^n-1} - \frac{(2^n-1)!}{1!(2^n-1)!}(1-1)^{2^n-1}\right) = \frac{1}{(2^n-2)!}.$$

Therefore, the density should be equal to the inverse of that quantity. ∎

Using the expression of the density, it is possible to prove that the measure of any subset follows a Beta distribution with known parameters.

Proposition 4.36 *[27] Let μ_{bel} be a uniform random belief measure. Then, for any $A \subset [n]$ with $A \neq \emptyset$, $\mu_{bel}(A) \sim Beta(2^{|A|}-1, 2^n - 2^{|A|})$.*

Proof: Using the transformation from the Mobius transform and the capacity, one can obtain

$$f_{\mu_{bel}(A)}(t) = f_{\sum_{B\subseteq A} M_{\mu_{bel}}(B)}(t) = f_{\sum_{i=1}^{2^{|A|}-1} X_i \,\middle|\, \sum_{i=1}^{2^n-1} X_i = 1}(t) =$$
$$= \frac{f_{\sum_{i=1}^{2^{|A|}-1} X_i, \sum_{i=1}^{2^n-1} X_i}(t,1)}{f_{\sum_{i=1}^{2^n-1} X_i}(1)} = \frac{f_{\sum_{i=1}^{2^{|A|}-1} X_i}(t) f_{\sum_{i=2^{|A|}}^{2^n-1} X_i}(1-t)}{f_{\sum_{i=1}^{2^n-1} X_i}(1)},$$

where $X_1, \ldots, X_{2^n-1}$ are independent standard uniform random variables.

The densities in the numerator can be computed by using Proposition 2.72:

$$f_{\sum_{i=1}^{2^{|A|}-1} X_i}(t) = \frac{1}{(2^{|A|}-2)!} \sum_{k=0}^{\lfloor t \rfloor} \frac{(2^{|A|}-1)! t^{2^{|A|}-2}}{k!(2^{|A|}-1-k)!} = \frac{t^{2^{|A|}-2}}{(2^{|A|}-2)!}$$
$$f_{\sum_{i=2^{|A|}}^{2^n-1} X_i}(1-t) = \frac{1}{(2^n - 2^{|A|}-1)!} \sum_{k=0}^{\lfloor 1-t \rfloor} \frac{(2^n-2^{|A|})!(1-t)^{2^n-2^{|A|}-1}}{k!(2^n-2^{|A|}-k)!} =$$
$$= \frac{(1-t)^{2^n-2^{|A|}-1}}{(2^n-2^{|A|}-1)!},$$

and the probability in the denominator, as stated in the proof of Proposition 4.35, is $\frac{1}{(2^n-2)!}$. Therefore, the density function is

$$f_{\mu_{bel}(A)}(t) = (2^n-2)! \frac{t^{2^{|A|}-2}}{(2^{|A|}-2)!} \frac{(1-t)^{2^n-2^{|A|}-1}}{(2^n-2^{|A|}-1)!} = \frac{t^{2^{|A|}-2}(1-t)^{2^n-2^{|A|}-1}}{B(2^{|A|}-1, 2^n-2^{|A|})},$$

where $B(\cdot,\cdot)$ is the Beta function. It is concluded that $\mu_{bel}(A) \sim Beta(2^{|A|}-1, 2^n - 2^{|A|})$. ∎

4.3.3.3 Possibility measures

Possibility measures can be expressed as the measure of the singletons (since the rest of the values can be obtained by computing the maximum). In addition, the measure of one of the singletons should be 1.

In this direction, the set of possibility measures can be expressed as $\{\vec{x} \in [0,1]^n \mid \max(\vec{x}) = 1\}$. The volume of this subset (in dimension $n-1$) is computed in the next result.

Lemma 4.37 *[27] The Lebesgue measure of dimension $n-1$ of*

$$S = \{\vec{x} \in [0,1]^n \mid \max(\vec{x}) = 1\},$$

is $\lambda(S) = n$ (with $n \geq 2$).

Proof: Express S as $S = \cup_{i=1}^{n} S_i$ with $S_i = \{\vec{x} \in [0,1]^n \mid x_i = 1\}$. Trivially, the $(n-1)$-Lebesgue measure of S_i is 1 for each $i \in [n]$. In addition, for each $i, j \in [n]$, $S_i \cap S_j = \{\vec{x} \in [0,1]^n \mid x_i = x_j = 1\}$, which has measure 0. Therefore, the measure of S is the sum of the measures of $S_1, \ldots, S_n$ and the result holds. ∎

Therefore, the density function of the measure of the singletons is

$$f_{\mu_p(\{i\}),\ldots,\mu_p(\{n\})}(\vec{x}) = \frac{1}{n}\mathbb{I}_S,$$

and the rest of measures are totally determined by the ones of the singletons. The following proposition simplifies future computations.

Proposition 4.38 *[27] A random possibility measure μ_p has uniform distribution if and only if the following properties are fulfilled:*

- *$P(\mu_p(\{i\}) = 1) = \frac{1}{n}$ for any $i \in [n]$,*
- *Given $\mu_p(\{i\}) = 1$, the distribution of $(\mu_p(\{j\}), j \neq i)$ is the same as $n-1$ independent uniform random variables.*

Proof: Starting with the necessity of the conditions,

- $P(\mu_p(\{i\}) = 1)$ is the volume of the set S_i defined in the proof of Lemma 4.37 divided by the volume of S, thus equals $\frac{1}{n}$ for each $i \in [n]$.
- Given $\mu_p(\{i\}) = 1$, the density function of $(\mu_p(\{j\}), j \neq i)$ can be computed as

$$f_{[(\mu_p(\{j\}),j\neq i) \mid \mu_p(\{i\})=1]}(\vec{x}) = \frac{\frac{1}{n}\mathbb{I}_{\vec{x}\in[0,1]^{n-1}}}{\frac{1}{n}} = \mathbb{I}_{\vec{x}\in[0,1]^{n-1}},$$

  which coincides with the distribution of $n-1$ independent uniform random variables.

For sufficiency, using the second point, the probability of the existence of a unique $i \in [n]$ such that $\mu(\{i\}) = 1$ is 1. Then, the density can be computed as follows.

$$f_{\mu_p(\{i\}),\ldots,\mu_p(\{n\})}(\vec{x}) = \sum_{i=1}^{n} P(\mu_p(\{i\}) = 1)\mathbb{I}_{\vec{x}\in[0,1]^n, x_i=1} =$$
$$= \sum_{i=1}^{n} \frac{1}{n}\mathbb{I}_{\vec{x}\in[0,1]^n, x_i=1} = \frac{1}{n}\mathbb{I}_S.$$

■

As a conclusion of the latter properties, it can be stated that the random variable that returns the index $i \in [n]$ for which $\mu(\{i\}) = 1$ is almost surely well defined and each index is equiprobable.

Corollary 4.39 *[27] Let μ_p be a uniform random possibility measure. Then, the events $E_1, \ldots, E_n$ with $E_i = \{\mu_p(\{i\}) = 1, \mu_p(\{j\}) \neq 1 \text{ if } j \neq i\}$, fulfill $P(E_i) = \frac{1}{n}$ and $P(E_i \cap E_j) = 0$ for any $i, j \in [n]$.*

Finally, it is possible to derive the distribution of the measures $\mu(A), A \subseteq [n]$.

Proposition 4.40 *[27] Let μ_p be a uniform random possibility measure of dimension n and $A \subseteq [n]$. Then, the distribution function of $\mu_b(A)$ has the following distribution function:*

$$F_{\mu_p(A)} = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{n-|A|}{n}x^{|A|} & \text{if } x \in (0,1), \\ 1 & \text{if } x \geq 1. \end{cases}$$

Proof: Since $\mu_p(A) = \max_{i \in A} \mu_p(\{i\})$, Corollary 4.39 implies that $P(\mu_p(A) = 1) = \sum_{i \in A} P(\mu_{\{i\}} = 1) = \frac{|A|}{n}$.

If $\mu_p(A) \neq 1$, applying point 2 in Proposition 4.38 and that any subset of variables of independent uniform random variables are also independent uniform random variables, it holds that $(\mu(\{i\}), i \in A)$ have the distribution of $|A|$ independent uniform random variables.

The distribution of the maximum of $|A|$ independent uniform random variables has as distribution function $F(x) = x^{|A|}$ if $x \in [0,1]$, see Proposition 2.83. This case occurs with probability $\frac{n-|A|}{n}$. The result holds by doing the mixture of the two cases. ∎

4.4 Stochastically Ordered Aggregation Operators

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 and the IOWA operators, as well as the Choquet and Sugeno integrals. 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 [331], the usual stochastic order is not a total order in the set of probability distributions, so pairs of incomparable random variables may appear.

This section is devoted to the definition of aggregations of random variables based on the stochastic ordering 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 the ordering.

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 [265]. A similar problem appears when dealing with random variables and stochastic orders. An alternative is to order the random variables in terms of their means and considering admissible permutations (see [264] for the definition of such permu-

tations). The approach does not have good properties in terms of monotonicity, although a slight modification of it has good properties when working with multivariate Gaussian random vectors. This alternative is not going to be disclosed here, the reader is referred to Section 3 in [39] for more information in this regard.

In the following sections, a second technique that transforms the initial random vector into a new one in which its components are ordered in the usual stochastic order will be used, following a similar approach to the one by Yager in [331]. Its main properties are studied, as well as the particular case of Stochastically Ordered Weighting Averagings, which will be of interest in the prediction problem presented in Section 7.4.

4.4.1 Definition of SOA operators

Stochastically Ordered Aggregations, SOA for short, will be a type of operator that rely on a transformation that generates a random vector from any given random vector, ensuring that its components are arranged according to the usual stochastic order.

Example 4.41 *[39] Consider the distribution functions of a uniform distribution over the interval* $[0,1]$*, an exponential with* $\lambda = 1$ *and a Chi-squared with 1 degree of freedom variables. On the left side of Figure 4.2, it can be seen that the lines intersect, indicating that they are not ordered according to the usual stochastic order. Ordering the distribution functions pointwise, three new distributions, which are ordered, appear as shown on the right side of Figure 4.2.*

In the following, starting with some distribution functions $F_1, \ldots, F_n$, it will be denoted as $F_{[1]}, \ldots, F_{[n]}$ the distribution functions obtained as the pointwise ordering of $F_1, \ldots, F_n$, being $F_{[1]}$ the greatest distribution function, which is associated with smaller values.

With this ordering, a set of ordered distribution functions is obtained. However, the objective is to work with random variables and their realizations, not with their distributions. In this direction, the variability of random variables will be considered by applying to them their distribution functions.

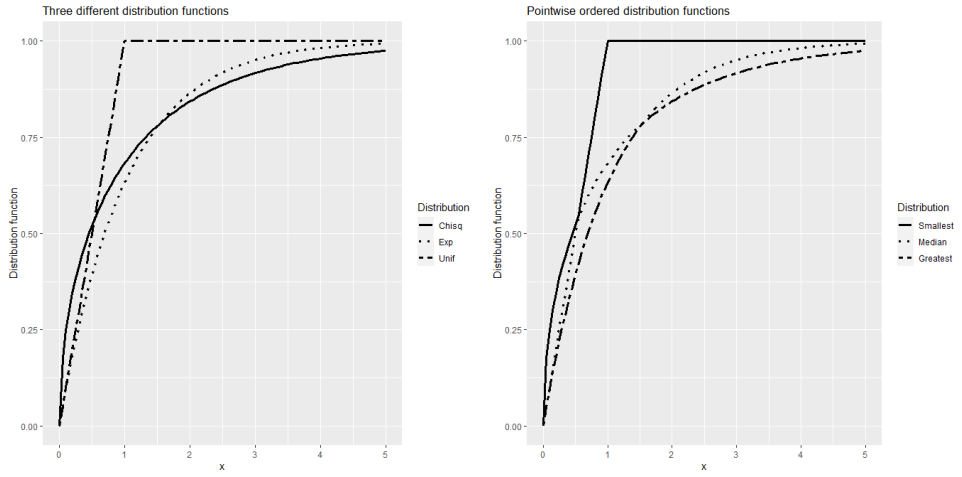

Figure 4.2: Three different cumulative distribution functions (left) and the cumulative distribution functions obtained by a pointwise ordering (right) [39].

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 standard uniform random variable (see [253]). Therefore, the variability of X can be totally explained by 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 uniform variables can be chosen. Moreover, choosing a uniform random variable will not allow conditional monotonicity (see Definition 3.47). A natural choice in this case can be, again, to consider $F_X(X)$ as a random variable that captures the variability of X.

In conclusion, given a set of random variables, it is possible to first build the ordered distributions $F_{[1]}, \ldots, F_{[n]}$ and then assign them the associated $F_1(X_1), \ldots, F_n(X_n)$ to obtain a new random vector that is stochastically ordered. The resulting random vector will be called the rearrangement increasing stochastically ordering of the initial one.

Definition 4.42 *[39] Let $\vec{X}$ be a random vector and $\pi : [n] \to [n]$ a permutation. Then, the rearrangement increasing stochastically ordered random vector, denoted as $\vec{X}_{so,\pi}$, is defined as:*

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

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

The semantics of the permutation π should be properly explained. If $\pi(i)=j$, the variability of X_j is associated with the distribution $F_{[i]}^{-1}$ (which is the i-th smallest). It allows $\vec{X}_{so,\pi}$ to be totally determined by $\vec{X}$ in the sense that the conditional distribution of $\vec{X}_{so,\pi}$ given a value of $\vec{X}$ is degenerate. This will be relevant in applied problems (see Section 7.4) 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,

Definition 4.43 *[39] Let I be a real interval, $A:I^n\to I$ a measurable aggregation function and $\pi:[n]\to[n]$ a permutation. Then, the Stochastically Ordered Aggregation is a function $SOA_{\hat{A},\pi}:L_I^n\to L_I$ such that for any $\vec{X}\in L_I^n$, $SOA_{\hat{A},\pi}(\vec{X})=A\circ\vec{X}_{so,\pi}$.*

Stochastically Ordered Aggregations can be seen as induced aggregations applied over rearrangement increasing stochastically ordered random vectors.

Example 4.44 *[39] Let $\vec{X}$ be a random vector with independent components such that $X_1\sim Exp(1.5)$ and $X_2\sim U[0,1]$. It is easy to see that $F_2(t)\leq F_1(t)$ 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 π such that $\pi(1)=2$ and $\pi(2)=1$ and $\hat{A}$ the arithmetic mean, $SOA_{\hat{A},\pi}(\vec{X})$ has the following expression.*

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

4.4.2 Monotonicity properties

Definition 4.43 introduces the SOA operators, but its relationship with the concept of aggregation of random variables, although it seems reasonable, must be proved. In particular, the main desirable property of the operator is monotonicity with respect to a stochastic order. Firstly, a preliminary result regarding the rearrangement increasing stochastically ordered random vector will be proved. For it, recall the definition of $\leq_{sd-st}$ provided in Definition 3.5.

Theorem 4.45 *[39] Let $\vec{X}$ and $\vec{Y}$ be two random vectors such that $\vec{X} \leq_{sd-st} \vec{Y}$ and let $\pi : [n] \to [n]$ be any permutation. Then, $\vec{X}_{so,\pi} \leq_{st} \vec{Y}_{so,\pi}$.*

Proof: Denote as $F_1, \ldots, F_n$ and $G_1, \ldots, G_n$ the marginal distribution functions of $\vec{X}$ and $\vec{Y}$. 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 $i \in [n]$, it holds that $F_i(t) \geq G_i(t)$ for any $i \in [n]$ and $t \in \mathbb{R}$. Then, it holds that $F_{[i]}(t) \geq G_{[i]}(t)$ for any $i \in [n]$ and $t \in \mathbb{R}$ and also $F^{-1}_{[i]}(t) \leq G^{-1}_{[i]}(t)$ for any $i \in [n]$ and $t \in [0,1]$.

As noted after Definition 3.5, $(F_1(X_1), \ldots, F_n(X_n))$ and $(G_1(Y_1), \ldots, G_n(Y_n))$ have the same distribution. Consider a random vector $\vec{Z}$ with the same distribution as $(F_1(X_1), \ldots, F_n(X_n))$ and $(G_1(Y_1), \ldots, G_n(Y_n))$. Then,

$$\begin{aligned}\vec{X}_{so,\pi} &=_{st} \left(F^{-1}_{[1]}\left(Z_{\pi(1)}\right), \ldots, F^{-1}_{[n]}\left(Z_{\pi(n)}\right)\right) \leq_{a.s.} \\ &\leq_{a.s.} \left(G^{-1}_{[1]}\left(Z_{\pi(1)}\right), \ldots, G^{-1}_{[n]}\left(Z_{\pi(n)}\right)\right) =_{st} \vec{Y}_{so,\pi}\,,\end{aligned}$$

which implies the usual stochastic order by using Theorem 2.102. ∎

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

Corollary 4.46 *[39] Any Stochastically Ordered Aggregation $SOA_{\hat{A},\pi}$ is an aggregation of random variables with respect to $\leq_{sd-st}$ for any aggregation function $\hat{A} : I^n \to I$ and any permutation $\pi : [n] \to [n]$.*

Proof: Let $\vec{X} \leq_{sd-st} \vec{Y}$. Using Theorem 4.45, one has that $\vec{X}_{so,\pi} \leq_{st} \vec{Y}_{so,\pi}$ for any permutation π. Since $SOA_{\hat{A},\pi}(\vec{X}) = A \circ \vec{X}_{so,\pi}$ and $SOA_{\hat{A},\pi}(\vec{Y}) = A \circ \vec{Y}_{so,\pi}$, applying

Theorem 3.12, it is concluded that $SOA_{\hat{A},\pi}(\vec{X}) \leq_{st} SOA_{\hat{A},\pi}(\vec{Y})$. The usual stochastic order and the same dependence structure usual stochastic order are equivalent for random variables, thus the monotonicity holds.

If the interval I is bounded, then the boundary conditions are straightforward to prove. If I does not have a lower bound, let $X \in L_I$. Then, using Theorem 3.12, there exists $\vec{X} \in L_I^n$ such that $\hat{A} \circ \vec{X} \leq_{st} X$. Now, consider a random vector $\vec{Y}$ that has the same copula as $\vec{X}$ and has marginal distribution functions $F_1, \ldots, F_n$ fulfilling $F_i = F_{[1]}$ for any $i \in [n]$. It is clear that $\vec{Y} =_{st} \vec{Y}_{so,\pi}$ for any permutation π and, using Proposition 2.103, that $\vec{Y} \leq_{st} \vec{X}$. Then, $A(\vec{Y}) =_{st} \hat{A} \circ \vec{Y}_{so,\pi} \leq_{st} \hat{A} \circ \vec{X} \leq_{st} X$. If there is not an upper bound for I, proceed analogously. ■

The latter result can be seen as the equivalent result of the Composition Theorem (Theorem 3.12) for Stocastically Ordered Aggregations. This property can be illustrated in the case of Example 4.44.

Example 4.47 *[39] Consider the same conditions as in Example 4.44 and let $\vec{Y}$ be a random vector with independent components such that $Y_1 \sim Exp(1)$ and $Y_2 \sim U[0.5, 1.5]$. Trivially, $\vec{X} \leq_{st} \vec{Y}$ and both random vectors have the same and unique copula, so $\vec{X} \leq_{sd-st} \vec{Y}$. If the distribution of $SOA_{\hat{A},\pi}(\vec{X})$ and $SOA_{\hat{A},\pi}(\vec{Y})$ is simulated, by plotting their empirical distributions one can observe that they are ordered, see Figure 4.3. Therefore, $SOA_{\hat{A},\pi}(\vec{X}) \leq_{st} SOA_{\hat{A},\pi}(\vec{Y})$.*

Another interesting monotonicity property that is especially relevant in applied problems (see Section 7.4), is the conditional monotonicity, introduced in Definition 3.47. Since distribution, quantile and aggregation functions are all increasing, this property is easy to prove for any SOA operator.

Proposition 4.48 *[39] Any Stochastically Ordered Aggregation $SOA_{\hat{A},\pi}$ is conditionally monotone.*

Proof: Consider a Stochastically Ordered Aggregation $SOA_{\hat{A},\pi} : L_I^n \to L_I$ and a random vector $\vec{X} \in L_I^n$. Then,

$$\left[SOA_{\hat{A},\pi}(\vec{X}) \;\middle|\; \vec{X} = \vec{x}\right] =_{a.s.} A\left(F_{[1]}^{-1}\left(F_{\pi(1)}(x_{\pi(1)})\right), \ldots, F_{[n]}^{-1}\left(F_{\pi(n)}(x_{\pi(n)})\right)\right),$$

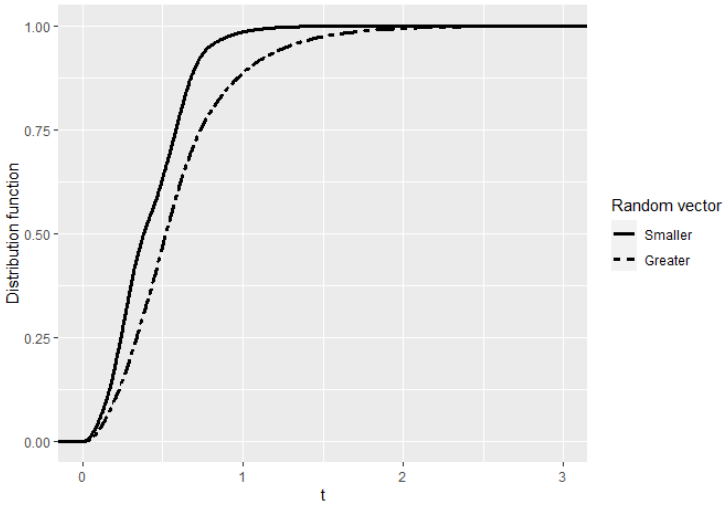

Figure 4.3: Simulated cumulative distribution functions of a Stochastically Ordered Aggregation of two ordered random vectors with sample size 10^6 [39].

for any $\vec{x} \in I^n$ such that the latter conditional distribution is well-defined.

From the latter expression, it is immediate that $[SOA_{\hat{A},\pi}(\vec{X}) \mid \vec{X} = \vec{x}]$ has degenerate distribution for any $\vec{x} \in I^n$ for which it is well-defined. Moreover, $[SOA_{\hat{A},\pi}(\vec{X}) \mid \vec{X} = \vec{x}] \leq [SOA_{\hat{A},\pi}(\vec{X}) \mid \vec{X} = \vec{y}]$ for any $\vec{x}, \vec{y} \in I^n$ such that $\vec{x} \leq \vec{y}$, since it is a composition of the increasing functions $F^{-1}_{[1]}, \ldots, F^{-1}_{[n]}, F_1, \ldots, F_n$ and $\hat{A}$ with a permutation of the random vector $\vec{X}$. ■

Similarly as in previous examples, the property can be illustrated for the same conditions.

Example 4.49 *[39] Consider the SOA operator considered in Example 4.44. Representing the value that the SOA operator takes with respect to the values of the input random variables, the surface represented in Figure 4.4 is obtained. 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 monotonicity properties, the following lemma links SOA operators applied to comonotone random variables with the composition of usual aggre-

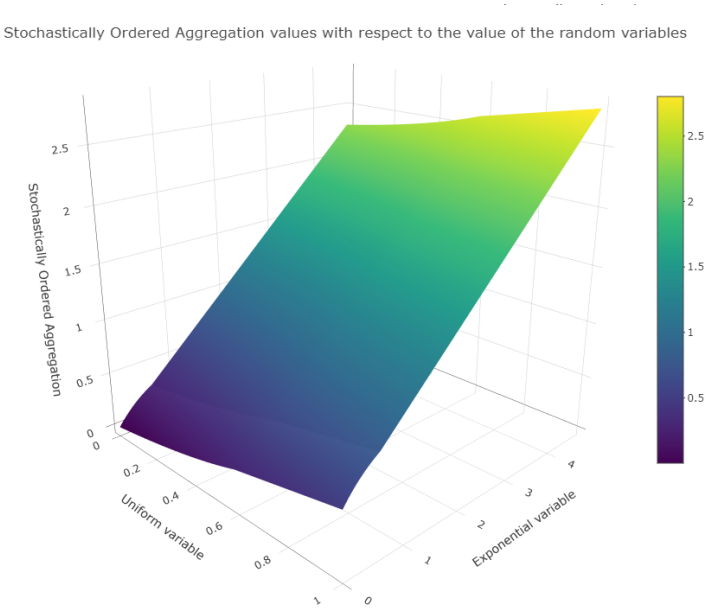

Figure 4.4: Representation of the value of the SOA operator versus the values of the input random variables in Example 4.44.

gation functions and order statistics.

Lemma 4.50 *[39] Let $\vec{X}$ be a random vector with continuous comonotone components. Then, $SOA_{\hat{A},\pi}(\vec{X}) =_{a.s.} \hat{A} \circ \left(X_{(1)},\ldots,X_{(n)}\right)$.*

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

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

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

$$\left(X_{(1)},\ldots,X_{(n)}\right) =_{a.s.} \left(F_{[1]}^{-1}(U),\ldots,F_{[n]}^{-1}(U)\right) =_{a.s.} \vec{X}_{so,\pi},$$

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

4.4.3 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 arithmetic mean. From a theoretical point of view, this particular choice is interesting because additional properties can be proved. From a practical one, it is a family of aggregations of random variables that are parametrized by a permutation and a weighting vector, which can be easily optimized when working with data (see Section 7.4). Its definition is a particular case of Definition 4.43.

Definition 4.51 *[39] Let $\vec{w} \in [0,1]^n$ be a weighting vector and $\pi : [n] \to [n]$ a permutation. The Stochastic Ordered Weighting Averaging (SOWA) operator, $SOWA_{\vec{w},\pi} : L_I^n \to L_I$, associated with $\vec{w}$ is defined as*

$$SOWA_{\vec{w},\pi}(\vec{X}) = \sum_{k=1}^{n} w_k \left(\vec{X}_{so,\pi}\right)_k,$$

for any $\vec{X} \in L_I^n$.

The weights of the SOWA operator have a relevant role, since they allow more or less importance to be given to each of the distributions $F_{[1]}, \ldots, F_{[n]}$. For instance, smaller weights can be given 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 central weights (see [140]), but instead of reducing the impact of extreme observations (outliers), it reduces the impact of extreme distributions.

In the following result some linearity properties of the SOWA operator are proved.

Proposition 4.52 *[39] Let $\vec{w} \in [0,1]^n$ be a weighting vector and $\pi : [n] \to [n]$ a permutation. Then,*

- *$SOWA_{\vec{w},\pi}(\lambda\vec{X}) = \lambda SOWA_{\vec{w},\pi}(\vec{X})$ for any random vector $\vec{X}$ and $\lambda \in \mathbb{R}^+$,*
- *$SOWA_{\vec{w},\pi}(\vec{X} + \lambda\vec{1}) = SOWA_{\vec{w},\pi}(\vec{X}) + \lambda$ for any random vector $\vec{X}$ and $\lambda \in \mathbb{R}$,*
- *$SOWA_{\vec{w},\pi}(\vec{X} + \vec{Y}) = SOWA_{\vec{w},\pi}(\vec{X}) + SOWA_{\vec{w},\pi}(\vec{Y})$ for any pair of random vectors with continuous components such that there exists a permutation $\hat{\pi}$ for*

which $X_{\hat{\pi}(1)} \leq_{st} \cdots \leq_{st} X_{\hat{\pi}(n)}$ *and* $Y_{\hat{\pi}(1)} \leq_{st} \cdots \leq_{st} Y_{\hat{\pi}(n)}$ *are satisfied and the random variables* X_i *and* Y_i *are comonotone for any* $i \in [n]$.

Proof: The first two statements are straightforward to prove from the definition of SOWA. For the third one, denote as $F_1, \ldots, F_n$, $G_1, \ldots, G_n$ and $H_1, \ldots, H_n$ the marginal 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 with $X_i + Y_i$ and Y_i with $X_i + Y_i$, are comonotone for any $i \in [n]$. 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 $i \in [n]$. Due to 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 $i \in [n]$.

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

$$\vec{U} =_{a.s} (F_1(X_1), \ldots, F_n(X_n)) =_{a.s.} (G_1(Y_1), \ldots, G_n(Y_n)) =_{a.s.} \\ =_{a.s.} (H_1(X_1 + Y_1), \ldots, H_n(X_n + Y_n)).$$

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

$$SOWA_{\vec{w},\pi}(\vec{X} + \vec{Y}) = \sum_{i=1}^{n} w_i H_{[i]}^{-1}\left(U_{\pi(i)}\right) = \\ = \sum_{i=1}^{n} w_i \left(F_{[i]}^{-1}\left(U_{\pi(i)}\right) + G_{[i]}^{-1}\left(U_{\pi(i)}\right)\right) = SOWA_{\vec{w},\pi}(\vec{X}) + SOWA_{\vec{w},\pi}(\vec{Y}).$$

■

Surprisingly, for continuous random vectors, its expected value is the same as the aggregation of random variables induced by an OWA operator.

Proposition 4.53 *[39] Let* $\vec{X}$ *be a random vector with continuous components. Then* $E[SOWA_{\vec{w},\pi}(\vec{X})] = \sum_{i=1}^{n} w_i E\left[X_{(i)}\right]$.

Proof: Noting that the components are continuous, consider $U_1, \ldots, U_n$ the standard uniform random variables such that $U_i = F_i(X_i)$ with $i \in [n]$. Then,

$$E[SOWA_{\vec{w},\pi}(\vec{X})] = E\left[\sum_{i=1}^{n} w_i F_{[i]}^{-1}(U_{\pi(i)})\right] = \sum_{i=1}^{n} w_i E\left[F_{[i]}^{-1}(U_{\pi(i)})\right] =$$
$$= \sum_{i=1}^{n} w_i \int_0^1 F_{[i]}^{-1}(t)dt = \int_0^1 \left(\sum_{i=1}^{n} w_i F_{[i]}^{-1}(t)\right) dt.$$

This expression does not depend on the copula of the random vector. The result holds since, using Lemma 4.50, for the case of comonotone components, it holds $SOWA_{\vec{w},\pi}(\vec{X}) =_{a.s.} \sum_{i=1}^{n} w_i X_{(i)}$. ∎

4.5 An approximation for aggregations of random variables

In recent papers (see [81] and [258]), a minimization problem related to the Gini mean difference [103] has been studied in detail. In particular, given a random variable X and a bivariate copula C, they look for a random variable Y such that the random vector (X,Y) has copula C and the quantity $E[|X-Y|]$ is as small as possible. The quantity $E[|X-Y|]$ can be seen as a distance measure between X and Y, while its minimum can be seen as a variability measure of X [82], being the Gini mean difference when Y is an independent copy of X. The solution of such a problem can be seen as the best approximation of a random variable by another one when the dependence between them is fixed, so the information that can be used from the first one in order to determine the second one is restricted.

In this section, a procedure to define aggregations of random variables by solving such a problem is provided. Firstly, the problem is solved in the general scenario, since in the aforementioned papers only partial solutions under some restrictions are given. Secondly, the monotonicity of the solutions of such a problem is proved, allowing one to define aggregation of random variables.

4.5.1 A solution to the general problem

As stated in [258], $E[|X-Y|]$ can be expressed as

$$E[|X-Y|] = \int_{-\infty}^{\infty} F(t) + G(t) - 2C(F(t), G(t))dt, \tag{4.6}$$

where F and G denote, respectively, the distribution function of X and Y.

In the aforementioned papers, the problem is restricted by considering that $G(t)$ should be of the form $h_\theta(F(t))$, with $(h_\theta, \theta \in \Theta)$ being a family of distortion functions. In addition, several assumptions are made for both C and X. However, in order to use the problem to define aggregations of random variables, the solution to the general problem should be stated.

In the following, considering a random variable X and a copula C, denote as $L_{X,C}$ the set of random variables such that, if $Y \in L_{X,C}$, then (X,Y) has copula C. The objective is to try to find $Y \in L_{X,C}$ such that

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|].$$

For proving the associated theorem, the notion of upper hemicontinuity and an associated result will be needed.

Definition 4.54 *[15] Given two topological spaces A and B, a function $H : A \to \mathscr{P}(B)$ is said to be upper hemicontinuous if for any $a \in A$ and V an open set such that $H(a) \subseteq V$, there exists a neighborhood of a, U, such that $H(a') \subseteq V$ for any $a' \in U$.*

The following result is a simplified version of Berge Maximum Theorem [54] that gives conditions for which, given a two-dimensional function, the set of maximizers in one of the variables is a non-empty, compact and upper hemicontinuous set-function with respect to the other variable.

Theorem 4.55 *[54] Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be a continuous function. Then, the set-valued function defined as $H : [0,1] \to \mathscr{P}([0,1])$ such that*

$$H(x) = \left\{ y \in [0,1] \;\middle|\; f(x,y) = \max_{z \in [0,1]} f(x,z) \right\} = \arg\max_{z \in [0,1]} f(x,z),$$

has non-empty compact values and is upper hemicontinuous.

The next theorem gives a solution for the general problem. No restrictions are considered for the distribution of X, the distribution of Y and the copula C. Minimizing the integrand point by point in Equation (4.6), a possible expression for $G(t)$ is obtained. Then, using the Berge Maximum Theorem and properties of copulas, it is possible to prove that the proposed $G(t)$ is indeed a distribution function.

Theorem 4.56 *[32] Let C be a copula. Then, given a random variable X, there exists a random variable Y such that (X,Y) has copula C and*

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|].$$

Proof: Start by noticing that any copula is a continuous function [106]. Denoting the distribution function of X and Y as F and G, one has

$$E\left[|X-Y|\right] = \int_{-\infty}^{\infty} F(t) + G(t) - 2C\left(F(t), G(t)\right) dt.$$

Since the values of $F(t)$ and $G(t)$ are in the interval $[0,1]$, it is possible to define the function $h : [0,1] \to [0,1]$ as

$$h(x) = \max\left(\arg\min_{y \in [0,1]} x + y - 2C(x,y)\right),$$

and then determine $G(t) = h(F(t))$. For this strategy to be valid, the only requirement that is needed is that $h(F(t))$ must be a distribution function for each possible distribution function $F(t)$. In particular, it is necessary to prove h to be well defined, increasing, right-continuous and that it fulfills $h(0) = 0$ and $\lim_{x \to 1} h(x) = 1$.

- h is well defined. Since C is continuous, $x + y + C(x,y)$ is also continuous. Apply Theorem 4.55 to $-(x+y+C(x,y))$. Then, $\arg\min_{y \in [0,1]} x+y-2C(x,y)$ is compact and non-empty for any $x \in [0,1]$. Then, the maximum exists and h is well defined.

- h is increasing. If h is not increasing, there exists $x \in [0,1]$ and $\varepsilon, \alpha > 0$ fulfilling $x + \varepsilon \leq 1$ and $h(x) - \alpha \geq 0$ such that $h(x+\varepsilon) = h(x) - \alpha$. Using the properties of copulas, see Definition 2.66, $C(x+\varepsilon, h(x)) - C(x, h(x)) - C(x+\varepsilon, h(x)-\alpha) + C(x, h(x)-\alpha) \geq 0$.

Therefore, one has

$$\begin{aligned} x+\varepsilon+h(x)-2C(x+\varepsilon,h(x))-x-\varepsilon-h(x)+\alpha+2C(x+\varepsilon,h(x)-\alpha)= \\ =x+h(x)-(x+h(x)-\alpha)-2\left(C(x+\varepsilon,h(x))-C(x+\varepsilon,h(x)-\alpha)\right)\leq \\ \leq x+h(x)-(x+h(x)-\alpha)-2\left(C(x,h(x))-C(x,h(x)-\alpha)\right)= \\ =x+h(x)-2C(x,h(x))-(x+h(x)-\alpha-2C(x,h(x)-\alpha))\leq 0, \end{aligned}$$

where the last equality is reached since $h(x)\in \arg\min_{y\in[0,1]}(x+y-2C(x,y))$. Then, $h(x)-\alpha \neq \max\left(\arg\min_{y\in[0,1]}(x+\varepsilon+y-2C(x+\varepsilon,y))\right)$ and therefore $h(x+\varepsilon)\neq h(x)-\alpha$. It is concluded that h is increasing.

- h is right continuous. Apply again Theorem 4.55 to obtain that the function $H:[0,1]\to\mathscr{P}([0,1])$ defined as $H(x)=\arg\min_{y\in[0,1]}x+y-C(x,y)$ has compact values and is upper hemicontinuous. In addition, there is the relation $h(x)=\max H(x)$.

  Using the same argument that for proving that h is increasing, it is proved that the values of H are connected, thus they are of the form $[a,b]\subseteq[0,1]$.

  Since h is increasing, for proving that h is right continuous it is enough to prove that for any $x\in[0,1)$ and $\varepsilon>0$ there exists $\delta>0$ such that $h(x+\delta)-h(x)<\varepsilon$.

  Consider the notation $H(x)=[a,b]$ and the open set $(a-\lambda,b+\lambda)$ with $\lambda<\min(a,1-b,\varepsilon)$. Then, using the upper hemicontinuity, there exists a neighborhood of x, U, such that for any $x'\in U$ one has $H(x')\subseteq(a-\lambda,b+\lambda)$. Then, for any $x'\in U$, $h(x')\leq b+\lambda$.

  Since U is a neighborhood of x, there exists $\delta_0>0$ such that for any $x'\in[0,1]$ such that $|x-x'|<\delta_0$, $x'\in U$. Therefore, considering $\delta_0=\delta$, one has $h(x+\delta)-h(x)\leq b+\lambda-b=\lambda<\varepsilon$. It is concluded that h is right continuous.

- $h(0)=0$. If $x=0$, then $0+h(0)-2C(0,h(0))=h(0)$ takes the minimum value if and only if $h(0)=0$.

- $\lim_{x\to1}h(x)=1$. Since $h(x)$ is increasing and bounded, the limit $\lim_{x\to1}h(x)$ exists. Suppose that $\lim_{x\to1}h(x)=x_0<1$. Using again that h is increasing,

one has that $h(x) \leq x_0$ for any $x \in [0,1]$. Now, let $x = 1-\delta$ with $\delta < \frac{1-x_0}{2}$. Then,

$$\begin{aligned} &1-\delta+h(1-\delta)-2C(1-\delta,h(1-\delta)) \geq \\ &\geq 1-\delta+h(1-\delta)-2C(1,h(1-\delta)) = \\ &= 1-\delta-h(1-\delta) > 1-\delta-x_0 > \frac{1-x_0}{2}. \end{aligned}$$

In addition, if $h(1-\delta)$ is replaced by 1,

$$1-\delta+1-2C(1-\delta,1) \geq 1-\delta+1-2(1-\delta) = \delta < \frac{1-x_0}{2}.$$

This is a contradiction to the fact that $h(1-\delta) < x_0$, since if its value is replaced for 1 the value of the objective function is smaller. It is concluded that $\lim_{x\to 1} h(x) = 1$. ∎

Notice that in the latter result not only the existence of a solution of the problem is proved but also a constructive method to determine the distribution function of a solution is given. However, the result does not say anything about the uniqueness of the solutions. Maybe another choice of a value in the set-valued function $H(x)$ can be made. Even with a unique distribution function, it is possible to define two random variables Y_1 and Y_2 that are not almost surely equal, fulfill $Y_1 =_{st} Y_2$ and such that (X,Y_1) and (X,Y_2) have the same copula.

4.5.2 Solutions for some particular copulas

In this section, the solution for some particular families of copulas is provided. For the comonotone case, the random variables X and Y have a perfect positive dependence. Then, trivially, the minimum of $E[|X-Y|]$ is reached when $X =_{a.s.} Y$. However, it is also possible to prove the statement using the introduced technique.

Proposition 4.57 *[32] Let C be the comonotone copula. Then, given a random variable X, for any random variable Y such that $X =_{a.s} Y$ one has that*

$$E[|X-Y|] = \min_{Z \in L_{X,C_{\min}}} E[|X-Z|].$$

Proof: The function that should be minimized is $x+y-2\min(x,y)$. Computing the derivative with respect to y, the result is -1 if $y<x$ and 1 if $y>x$. Therefore, since the function is continuous, the minimum is reached when $y=x$. Then, consider $h(x)=x$, thus Y has the same distribution as X and it is concluded that $X =_{a.s} Y$. ■

However, for the other two main copulas, the solution is not that straightforward. In particular, the best choice is to consider a degenerate random variable that takes the median of X as its value.

Proposition 4.58 *[32] Let C be the independence or the countermonotone copula. Then, given a random variable X with distribution function $F(t)$, for any random variable Y such that $Y =_{a.s.} F^{-1}(0.5)$, one has that*

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|].$$

Proof: For the independence copula, the function that should be minimized is $x+y-2xy$. Computing the derivative with respect to y, the result is $1-2x$, which is positive when $x<0.5$, negative when $x>0.5$ and 0 if $x=0.5$. Then, h can be considered to be $h(x)=0$ if $x<0.5$ and $h(x)=1$ otherwise. Therefore, $h(F(t))$ corresponds to a degenerate random variable with value $F^{-1}(0.5)$.

For the countermonotone copula, the function that should be minimized is $x+y-2\max(x+y-1,0)$. Computing the derivative with respect to y, the result is 1 if $x+y<1$ and -1 otherwise. Then, the minimum is reached in $y=0$ or in $y=1$. In the first case, one has that the function is $x+0-2\max(x+0-1,0)=x$ and in the second one $x+1-2\max(x+1-1,0)=1-x$, thus the minimum is reached in $y=0$ when $x<0.5$, in $y=1$ when $x<0.5$ and at both ends when $x=0.5$. For the rest of the steps, proceed as in the previous case. ■

Since the median is the solution for the independence and the countermonotone copula, a natural question is to consider if this also holds for any copula between them. In particular, since the countermonotone copula is the smallest, this is equivalent to proving the following result.

Proposition 4.59 *[32] Let C be a copula such that $C(x,y) \leq xy$ for all $x,y \in [0,1]$. Then, given a random variable X with distribution function $F(t)$, for any random*

variable Y such that $Y =_{a.s.} F^{-1}(0.5)$ one has that

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|].$$

Proof: Denote as C_1 and C_2 the independent and countermonotone copulas. Since C_2 is the smallest copula, one has $C_2(x,y) \leq C(x,y) \leq C_1(x,y)$ for any $x,y \in [0,1]$. Then, C can be expressed as $C(x,y) = \lambda(x,y)C_1(x,y) + (1-\lambda(x,y))C_2(x,y)$ with $\lambda(x,y) \in [0,1]$ for any $x,y \in [0,1]$. Then, the function to minimize is of the form

$$\begin{aligned} x+y-2C(x,y) &= x+y-2\left(\lambda(x,y)C_1(x,y) + (1-\lambda(x,y))C_2(x,y)\right) = \\ &= \lambda(x,y)\left(x+y-C_1(x,y)\right) + (1-\lambda(x,y))\left(x+y-C_2(x,y)\right). \end{aligned}$$

Using Proposition 4.58, the minimum of $(x+y-C_1(x,y))$ and the minimum $(x+y-C_2(x,y))$ for any $x \in [0,1]$ is reached at the same point, $y=0$ if $x<0.5$ and $y=1$ otherwise. Then, regardless of the value of $\lambda(x,y)$, the minimum of $x+y-2C(x,y)$ is also reached in $y=0$ when $x<0.5$ and $y=1$ otherwise. Then, the result is reached by defining h as in the proof of Proposition 4.58. ∎

In addition, a similar result is provided for FGM copulas (see Example 2.68), is provided.

Proposition 4.60 *[32] Let C be an FGM copula with parameter $\lambda \in [-1,1]$. Then, given a random variable X with distribution function $F(t)$, for any random variable Y such that,*

- $Y =_{a.s.} F^{-1}(0.5)$ *if* $\lambda \leq 0$,
- *Its distribution function is $h(F(t))$ with $h(x)$ being the median value between* 0, $\frac{1}{2}\frac{2\lambda x(1-x)-1+2x}{2\lambda x(1-x)}$ *and* 1 *(when $x \notin \{0,1\}$) if* $\lambda > 0$,

$$E[|X-Y|] = \min_{Z \in L_{X,C}} E[|X-Z|].$$

Proof: For $\lambda \leq 0$, the solution is given by Proposition 4.59.

Suppose that $\lambda > 0$. The function that should be minimized is $x+y-2xy-2\lambda x(1-x)y(1-y)$. Computing the derivative with respect to y, the result is $1-$

$2x-2\lambda x(1-x)(1-2y)$. Moreover, if the second derivative is computed, the result is $4\lambda x(1-x)$, which is positive for any $x \in [0,1]$.

The solution (for y) of $1-2x-2\lambda x(1-x)(1-2y)=0$ is $y=\frac{1}{2}\frac{2\lambda x(1-x)-1+2x}{2\lambda x(1-x)}$. If $y \in [0,1]$, then it is the minimum. If $y<0$, since the second derivative is positive, the derivative is positive in $[0,1]$ and therefore the minimum over $[0,1]$ is located in 0. Similarly, if $y>1$, the minimum over $[0,1]$ is located in 1. ∎

4.5.3 Approximations of aggregations of random variables

Recall again that the proof of Theorem 4.56 is constructive. In particular, if Y is a possible solution of the construction problem, it has distribution function $h(F(t))$, where $F(t)$ is the distribution function of X and h is defined as the maximum of the compact set of minimizers of the integrand (point by point) in the objective function, (see Equation 4.6). In addition, in that result, it is proved that the function h depends only on the copula and it is increasing. Using these properties, a monotonicity result regarding the solutions of the problem can be stated.

Proposition 4.61 *[32] Consider the same conditions as in Theorem 4.56. If X_1 and X_2 are two random variables such that $X_1 \leq_{st} X_2$, then there exist two random variables Y_1 and Y_2 such that:*

$$E[|X_1-Y_1|] = \min_{Z \in L_{X_1,C}} E[|X_1-Z|],$$

$$E[|X_2-Y_2|] = \min_{Z \in L_{X_2,C}} E[|X_2-Z|],$$

and $Y_1 \leq_{st} Y_2$.

Proof: Denote as $F_1(t)$ and $F_2(t)$ the distribution functions of X_1 and X_2. Since $X_1 \leq_{st} X_2$, $F_1(t) \geq F_2(t)$ for any $t \in \mathbb{R}$. As proved in Theorem 4.56, the distribution functions of Y_1 and Y_2 can be chosen to be $h(F_1(t))$ and $h(F_2(t))$ with h being increasing. Then, one has $h(F_1(t)) \geq h(F_2(t))$ for any $t \in \mathbb{R}$ and it is concluded that $Y_1 \leq_{st} Y_2$. ∎

The last result can be used to define aggregations of random variables considering bounded intervals. In particular, given an aggregation of random variables A,

any function constructed by solving the minimization problem, for a fixed copula and for each of $A(\vec{X})$ with $\vec{X} \in L_I$, is also an aggregation of random variables.

Corollary 4.62 *Let I be a bounded real interval and let $A : L_I^n \to L_I$ be an aggregation of random variables. Then, there exists an aggregation of random variables $B : L_I^n \to L_I$ such that*

$$E[|A(\vec{X}) - B(\vec{X})|] = \min_{Z \in L_{A(\vec{X}),C}} E[|A(\vec{X}) - Z|],$$

for any $\vec{X} \in L_I^n$.

Proof: Consider the function h defined in Theorem 4.56 and consider, for any $\vec{X} \in L_I^n$, the random variable $Y_{\vec{X}}$ with distribution function $h\left(F_{A(\vec{X})}\right)$ such that the random vector $\left(A(\vec{X}), Y_{\vec{X}}\right)$ has copula C. Then, applying Theorem 4.56, it holds

$$E[|A(\vec{X}) - Y_{\vec{X}}|] = \min_{Z \in L_{A(\vec{X}),C}} E[|A(\vec{X}) - Z|],$$

for any $\vec{X} \in L_I^n$.

For any $\vec{X} \in L_I^n$, define $S_{\vec{X}} = \{\omega \in \Omega \mid Y_{\vec{X}}(\omega) \in I\}$. If $S_{\vec{X}}$ has probability 1, then there exists a random variable $Z_{\vec{X}}$ such that $Y_{\vec{X}} =_{a.s.} Z_{\vec{X}}$ and $Z_{\vec{X}} \in L_I$. If $P(Y_{\vec{X}} < \inf I) > 0$ one can consider another random variable $Y'_{\vec{X}} \in L_I$ defined as $Y'_{\vec{X}}(\omega) = Y_{\vec{X}}(\omega)$ if $Y_{\vec{X}} \in I$ and $Y'_{\vec{X}}(\omega) = \inf I$ otherwise. Then, consider the random variable $Y''_{\vec{X}}$ such that $Y'_{\vec{X}} =_{st} Y''_{\vec{X}}$ and $(Y''_{\vec{X}}, A(\vec{X}))$ has copula C. It is clear that $E[|A(\vec{X}) - Y_{\vec{X}}|] > E[|A(\vec{X}) - Y''_{\vec{X}}|]$ and $Y^2_{\vec{X}} \in L_I$, which is a contradiction. If $P(B(\vec{X}) < \sup I) > 0$, proceed analogously.

Then, $S_{\vec{X}}$ has probability 1 for any $\vec{X} \in L_I^n$ and it is possible to define $B : L_I^n \to L_I$ as $B(\vec{X}) = Z_{\vec{X}}$ for any $\vec{X} \in L_I^n$. This function, since $Z_{\vec{X}} \in L_I$ for any $\vec{X} \in L_I^n$, is well defined. It remains to prove the monotonicity and the boundary conditions of B. The monotonicity is a direct consequence of Proposition 4.61 and the monotonicity of A. Since the interval is bounded, it is clear that $A(a,\ldots,a) =_{st} B(a,\ldots,a) =_{st} a$ and $A(b,\ldots,b) =_{st} B(b,\ldots,b) =_{st} b$ and the boundary conditions hold. ∎

This result cannot be held for unbounded intervals. For instance, notice that if C is the independent copula, as a consequence of Proposition 4.58, the resulting

aggregation of random variables would be degenerate, which contradicts Proposition 3.15.

The construction method explained above could be useful to model scenarios in which the inputs and outputs have a weak dependence. For instance, consider a stationary multivariate time series $(\vec{X}_n, n \in \mathbb{N})$ and $\hat{A}$ an aggregation function. If the quantity of $\hat{A}(\vec{X}_n)$ is of interest, a simple prediction problem could be to try to predict such a quantity by the value of $\hat{A}(\vec{X}_{n-k})$, that is, to use the value associated with a previous time. In this direction, if $(\hat{A}(\vec{X}_{n-k}), \hat{A}(\vec{X}_n))$ has copula C, then one could use Theorem 4.56 to find a solution Y of the minimization problem with $X = \hat{A}(\vec{X}_n)$ and copula C. Notice that, as a consequence of stationarity, both $\hat{A}(\vec{X}_{n-k})$ and $\hat{A}(\vec{X}_n)$ have the same distribution function F. Therefore, a solution can be computed as $Y = (h \circ F)^{-1}(F(\hat{A}(\vec{X}_{n-k})))$, with h as in Theorem 4.56 for any $n, k \in \mathbb{N}$ such that $k < n$. Notice that, since F and h are increasing, the function applied to $\hat{A}(\vec{X}_{n-k})$ is increasing and, therefore, the copula of $(Y, \hat{A}(\vec{X}_n))$ is C.

Chapter 5

Stochastic inequalities

Contents

The probabilistic approach developed in Chapter 3 not only allows new ways to aggregate random data. Even for usual aggregation functions, working with random variables makes it possible to study some properties in more detail. This fact is not new in the literature, Chapter 10 in [149] is devoted to the *behavioral analysis of aggregation functions*. This behavioral analysis consists of assuming that the inputs of the aggregation functions follow a particular distribution and then compute a quantity of interest, such as its mean. A similar approach, considering uniform distributions, is used in [267] to define a measure for aggregation functions.

The objective of this chapter is to give some stochastic inequalities that describe the behavior of aggregation functions and related functions in probabilistic terms. It is important to remark that this type of results can be proven only when considering a probabilistic approach. It is not possible to work with concepts such as variability or dependence without randomness.

In particular, the study will focus on the property of reducing the variability of some aggregation functions and the location comparison of some variability measures that are related to aggregation functions based on penalties [78]. For the first

case, Section 5.1, some cases in which the output of induced aggregations of random variables is smaller with respect to the convex stochastic order than the inputs are provided. These results formally prove the intuition that, when some aggregation functions are applied, the variability is reduced. For the second case, variability measures are ordered by location stochastic orders considering that the initial random vectors are ordered with respect to a variability stochastic order (Section 5.2) or by a dependence stochastic order (Section 5.3). These results can be applied to study the behavior of the minimum of penalty functions and also to define some hypothesis tests (see Section 7.5).

5.1 Reduction of variability

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with finite variance. It is a well-known fact that

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\mathrm{Var}(X_1)}{n},$$

so the application of the arithmetic mean induces a reduction of the variance with respect to the original distribution of each of the random variables $X_1, \ldots, X_n$. However, it is not possible to find results of this kind in the literature dealing with:

i) Stronger notions of comparisons in terms of variability or dispersion,

ii) Non-independent random variables,

iii) More general families of mean functions.

In this section, formal proofs of intuitions such as, for example, the fact that averages of central order statistics have less variability than averages of the extreme ones are provided. On this aim, results describing conditions for a random vector $\vec{X}$ and a mean function f in order to satisfy

$$f(\vec{X}) \leq_{cx} X_j, \forall j \in [n],$$

are proved. Recall that $\leq_{cx}$ denotes the convex order, see Definition 2.116. In particular, the results provided deal with families of weighted quasi-arithmetic means,

OWA operators, nullnorms and uninorms. To prove such statements, the following lemma will be used.

Lemma 5.1 *[41] Let h_1 and h_2 be two centrally symmetric distortions such that $h_1(t) \leq h_2(t)$ for any $t \in [0,0.5]$, and let X be a symmetric random variable with distribution function F. Denote as X_1 and X_2 two random variables with distribution functions $h_1(F)$ and $h_2(F)$, respectively. Then,*

$$X_1 \leq_{cx} X_2.$$

Proof: Firstly, it is clear that, since h_1 and h_2 are centrally symmetric and X is symmetric, both X_1 and X_2 are symmetric and have the same median m. Therefore, if it exists, they also have the same mean m.

Consider the integrals

$$I_1(x) = \int_{-\infty}^{x} h_1(F(t))dt \quad \text{and} \quad I_2(x) = \int_{-\infty}^{x} h_2(F(t))dt.$$

Let $x \leq m$. In this case, $h_1(F(t)) \leq h_2(F(t))$ for any $t \in (-\infty, x]$, and therefore $I_1(x) \leq I_2(x)$ holds for any $x \leq m$.

Let $x \geq m$. Note that, since X is symmetric, for all $t \geq m$ it holds $F(t) = \bar{F}(2m-t)$. Moreover, since h_1 is centrally symmetric, one has $h_1(F(t)) = h_1(\bar{F}(2m-t)) = h_1(1-F(2m-t)) = 1-h_1(F(2m-t))$. It follows that

$$I_1(x) = \int_{-\infty}^{x} h_1(F(t))dt = \int_{-\infty}^{m} h_1(F(t))dt + \int_{m}^{x} h_1(F(t))dt$$
$$= \int_{-\infty}^{m} h_1(F(t))dt + \int_{m}^{x} (1-h_1(F(2m-t)))\,dt = \int_{-\infty}^{2m-x} h_1(F(t))dt + (x-m).$$

Similarly, one has that

$$I_2(x) = \int_{-\infty}^{2m-x} h_1(F(t))dt + (x-m).$$

Observing now that, since $x \geq m$, $2m-x \leq m$, it follows that $I_1(x) \leq I_2(x)$ holds in this case as well. It is concluded that $I_1(x) \leq I_2(x)$ for any $x \in \mathbb{R}$ and, as a consequence of Theorem 2.117 and $E[X_1] = E[X_2]$, that $X_1 \leq_{cx} X_2$. ∎

5.1.1 Reduction of variability of weighted quasi-arithmetic means

In this section, the case of weighted quasi-arithmetic means is considered. As introduced in Definition 2.15, there is a linear convex combination involved in their computation, so one may expect a good behavior with respect to the convex order. In fact, this is the case for weighted arithmetic means.

Theorem 5.2 *[41] Let $\vec{X}$ be a random vector of dimension n with identically distributed components and $\vec{w} \in \mathbb{R}^n$ a weighting vector. Then,*

$$\sum_{i=1}^{n} w_i X_i \leq_{cx} X_j,$$

for any $j \in [n]$.

Proof: Let $\phi : \mathbb{R} \to \mathbb{R}$ be any convex function. Then,

$$E\left[\phi\left(\sum_{i=1}^{n} w_i X_i\right)\right] \leq E\left[\sum_{i=1}^{n} w_i \phi(X_i)\right] = \sum_{i=1}^{n} w_i E[\phi(X_i)].$$

Since $X_1, \ldots, X_n$ have the same distribution, it holds that $\phi(X_1), \ldots, \phi(X_n)$ also have the same distribution. Therefore,

$$E\left[\phi\left(\sum_{i=1}^{n} w_i X_i\right)\right] \leq \sum_{i=1}^{n} w_i E[\phi(X_i)] = \sum_{i=1}^{n} w_i E[\phi(X_j)] = E[\phi(X_j)],$$

for any $j \in [n]$. ∎

Notice that the result holds even in the case of non-independent components. However, it cannot be extended to other weighted quasi-arithmetic means, mainly because the same expectation for the compared variables is a necessary condition for the convex order to hold. A simple example can be given by considering the geometric mean.

Example 5.3 *[41] Let (X_1, X_2) be a random vector with standard uniform and independent components. Then the inequality $\sqrt{X_1 X_2} \leq_{cx} X_1$ cannot be true, since the two random variables have different expectations,*

$$E[\sqrt{X_1 X_2}] = E[\sqrt{X_1}]E[\sqrt{X_2}] = \frac{4}{9} \neq \frac{1}{2} = E[X_1].$$

Fortunately, if an adequate transformation is applied to both sides, one can prove a statement similar to Theorem 5.2.

Corollary 5.4 *[41] Let $f:\mathbb{R}^n \to \mathbb{R}$ be a weighted quasi-arithmetic mean with associated real-valued strictly increasing function h. Then, for any random vector $\vec{X}$ with identically distributed components,*

$$h(f(\vec{X})) \leq_{cx} h(X_j), \forall j \in [n].$$

Proof: Notice that $h(f(\vec{X})) = h(h^{-1}(\sum_{i=1}^n w_i h(X_i))) = \sum_{i=1}^n w_i h(X_i)$ and apply Theorem 5.2. ∎

Notice that the latter result can be applied for example, to the geometric, harmonic and power means.

5.1.2 Reduction of variability of OWA operators

In this section, Ordered Weighted Averaging operators consisting of independent and identically distributed random variables are considered. Recall that the distribution function of the *i*-th order statistic of a random vector with independent and identically components is

$$F_{(i)}(t) = \sum_{k=i}^{n} \binom{n}{k} F(t)^k (1-F(t))^{n-k},$$

where $F(t)$ is the distribution function of the components of $\vec{X}$, see Proposition 2.83. Therefore, it can be expressed as $F_{(i)}(t) = h_{(i)}(F(t))$, being

$$h_{(i)}(t) = \sum_{k=i}^{n} \binom{n}{k} t^k (1-t)^{n-k}.$$

Firstly, it is possible to prove a statement showing that the average of the distribution functions is smaller in the convex order for central order statistics than for extreme order statistics. Notice that this average is associated with a mixture of order statistics.

Lemma 5.5 *[41] Let $\vec{X}$ be a random vector with independent, symmetric and identically distributed components. Let Y_i be a random variable with distribution function $G_i(t) = \frac{1}{2}F_{(i)}(t) + \frac{1}{2}F_{(n-i+1)}(t)$ for any $t \in \mathbb{R}$. Then,*

$$|n-2i+1| \leq |n-2j+1| \implies Y_i \leq_{cx} Y_j,$$

for any $i, j \in [n]$

Proof: Consider $i \leq n-i+1$ and $j \leq n-j+1$. For the other cases, notice that $Y_j =_{st} Y_{n-j+1}$. Since $|n-2i+1| \leq |n-2j+1|$, then $j \leq i$.

The distortions associated with Y_i and Y_j can be written as

$$h_i(t) = \frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}t^k(1-t)^{n-k} + \frac{1}{2}\sum_{k=n-i+1}^{n}\binom{n}{k}t^k(1-t)^{n-k},$$

$$h_j(t) = \frac{1}{2}\sum_{k=j}^{n}\binom{n}{k}t^k(1-t)^{n-k} + \frac{1}{2}\sum_{k=n-j+1}^{n}\binom{n}{k}t^k(1-t)^{n-k}.$$

It can be observed that h_i is centrally symmetric, being

$$\begin{aligned}
h_i(1-t)+h_i(t) &= \frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}(1-t)^k t^{n-k} + \frac{1}{2}\sum_{k=n-i+1}^{n}\binom{n}{k}(1-t)^k t^{n-k} + \\
&+\frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}t^k(1-t)^{n-k} + \frac{1}{2}\sum_{k=n-i+1}^{n}\binom{n}{k}t^k(1-t)^{n-k} = \\
&= \frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}(1-t)^k t^{n-k} + \frac{1}{2} - \frac{1}{2}\sum_{k=0}^{n-i}\binom{n}{k}(1-t)^k t^{n-k} + \\
&+\frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}t^k(1-t)^{n-k} + \frac{1}{2} - \frac{1}{2}\sum_{k=0}^{n-i}\binom{n}{k}t^k(1-t)^{n-k} = \\
&= \frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}(1-t)^k t^{n-k} + \frac{1}{2} - \frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}t^k(1-t)^{n-k} + \\
&+\frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}t^k(1-t)^{n-k} + \frac{1}{2} - \frac{1}{2}\sum_{k=i}^{n}\binom{n}{k}(1-t)^k t^{n-k} = 1.
\end{aligned}$$

Similarly, the distortion h_j is also centrally symmetric. In addition, the difference between the distortions h_i and h_j is given by

$$h_i(t) - h_j(t) = \frac{1}{2} \sum_{k=n-i+1}^{n-j} \binom{n}{k} t^k (1-t)^{n-k} - \frac{1}{2} \sum_{k=j}^{i-1} \binom{n}{k} t^k (1-t)^{n-k}.$$

Notice that the index of the first summand can be reversed by the operation $n-k$, resulting in a unique sum over the same indices. Then,

$$h_i(t) - h_j(t) = \sum_{k=j}^{i-1} \binom{n}{k} \left(t^{n-k}(1-t)^k - t^k(1-t)^{n-k} \right).$$

Since k is at most $i-1$, $n-k$ is always greater than k. Therefore, if $t < 0.5$ all the summands are negative and the sum is negative. It is concluded that $h_1(t) \leq h_2(t)$ for any $t \in [0, 0.5]$. The statement now holds by Lemma 5.1. ∎

The latter result can be interpreted, when considering uniform mixtures, as the central order statistics having less variability than the extreme ones. Intuitively, one may think that, therefore, OWA operators having greater central weights should reduce the dispersion. In the next result, the intuition is formalized.

Theorem 5.6 *[41] Let $\vec{X}$ be a random vector with independent, symmetric and identically distributed components, and let $\vec{w} \in [0,1]^n$ be a weighting vector such that:*

i) $|n-2i+1| \leq |n-2j+1| \implies w_i \geq w_j$,

ii) $w_i = w_{n-i+1}$ *for any* $i \in [n]$.

Then,

$$\sum_{i=1}^{n} w_i X_{(i)} \leq_{cx} X_j,$$

for any $j \in [n]$.

Proof: Denote the distribution function of the components of $\vec{X}$ as F and consider $G_j(t) = \frac{1}{2}F_{(j)}(t) + \frac{1}{2}F_{(n-j+1)}(t)$ for any $t \in \mathbb{R}$, where $F_{(j)}$ and $F_{(n-j+1)}$ denote,

respectively, the distribution functions of $X_{(j)}$ and $X_{(n-j+1)}$. Notice that any component of $\vec{X}$ can be written as a mixture of the order statistics of $\vec{X}$, i.e.,

$$F(t) = \frac{1}{n}\sum_{i=1}^{n} F_{(i)}(t) = \frac{1}{n}\sum_{i=1}^{n} G_i(t),\ \forall t \in \mathbb{R}. \tag{5.1}$$

Consider now a random variable Z with distribution function F_Z defined as

$$F_Z(t) = \sum_{i=1}^{n} w_i F_{(i)}(t) = \sum_{i=1}^{n} w_i G_i(t), \tag{5.2}$$

for any $t \in \mathbb{R}$, where the second equality holds from the fact that $w_i = w_{n-i+1}$ for any $i \in [n]$.

Thus, the variable Z and any component of $\vec{X}$ can be written as a mixture of the random variables Y_i defined as before, which are in turn mixtures of the order statistics. By applying Lemma 5.5 one has that $|n-2i+1| \leq |n-2j+1|$ implies $Y_i \leq_{cx} Y_j$. In addition, the weights fulfill $|n-2i+1| \leq |n-2j+1| \implies w_i \geq w_j$. Then, for any convex function $\phi : \mathbb{R} \to \mathbb{R}$,

$$E[\phi(Z)] = \sum_{i=1}^{n} w_i E[\phi(Y_i)] \leq \frac{1}{n}\sum_{i=1}^{n} E[\phi(Y_i)] = E[\phi(X_j)], \forall j \in [n],$$

where the first equality holds by Equation (5.2), the second equality by the fact that $E[\phi(Y_i)] \leq E[\phi(Y_j)]$ if $w_i \geq w_j$ and the third one by Equation (5.1). Therefore, $Z \leq_{cx} X_j$ holds for any $j \in [n]$. Now observe that, again considering any convex function $\phi : \mathbb{R} \to \mathbb{R}$, it holds

$$E\left[\phi\left(\sum_{i=1}^{n} w_i X_{(i)}\right)\right] \leq E\left[\sum_{i=1}^{n} w_i \phi\left(X_{(i)}\right)\right] = \sum_{i=1}^{n} w_i E\left[\phi\left(X_{(i)}\right)\right] = E[\phi(Z)],$$

where the first inequality follows from the convexity of ϕ and the third one by Equation (5.2). It is concluded that $\sum_{i=1}^{n} w_i X_{(i)} \leq_{cx} Z \leq_{cx} X_j$, $j \in [n]$. ∎

Notice that Condition i) imposes that the weights should be greater for central positions, while Condition ii) refers to the symmetry of the weights. In the next example, the latter result is illustrated for the case of the median of three independent standard uniform random variables.

Example 5.7 *[41] Let $\vec{X} = (X_1, X_2, X_3)$ be a random vector with standard uniform components. Compute the integral of the distribution function of X_1 and $X_{(2)}$,*

$$\int_0^x F_1(t)dt = \int_0^x t\,dt = \frac{1}{2}x^2, \quad \int_0^x F_{(2)}(t)dt = \int_0^x \left(3t^2(1-t)+t^3\right)dt = x^3 - \frac{1}{2}x^4,$$

for any $x \in [0,1]$. Observing that $\frac{x^2}{2} \geq x^2 - \frac{x^4}{2}$ for any $x \in [0,1]$ and that $E[X_1] = E[X_{(2)}] = 0.5$, it is concluded that $X_1 \geq_{cx} X_{(2)}$.

The required symmetry for the distribution and the weights is due to the necessity of having the same expectation for the convex order to hold. Having symmetric and bigger central weights is a condition that appears naturally in some context of mean estimation of particular location-scale families related to, for example, Laplace, logistic or hyperbolic secant distribution (see Figure 7.1 for a representation of the associated cumulative weights) and also constitutes a wide family of OWA operators [329]. Unfortunately, if the assumption is removed, the result is not longer true, as shown in the following example.

Example 5.8 *[41] Let X_1, X_2 and X_3 be three independent and identically distributed random variables such that $P(X_i = 0) = P(X_i = 2) = 0.1$ and $P(X_i = 1) = 0.8$ for any $i \in [3]$. Consider $Y = \frac{1}{2}X_{(1)} + \frac{1}{2}X_{(3)}$. The values of Y and the associated probability for any possible value of (X_1, X_2, X_3) can be found in Tables 5.1 and 5.2.*

X_1 \ (X_2, X_3)	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$	$(2,0)$	$(0,2)$	$(2,1)$	$(1,2)$	$(2,2)$
0	0	0.5	0.5	0.5	1	1	1	1	1
1	0.5	0.5	0.5	1	1	1	1.5	1.5	1.5
2	1	1	1	1.5	1	1	1.5	1.5	2

Table 5.1: Values of $\frac{1}{2}X_{(1)} + \frac{1}{2}X_{(3)}$ associated with the possible values of the random vector (X_1, X_2, X_3) [41].

Therefore, summing all the probabilities for each value of Y, it holds that $P(Y = 0) = P(Y = 2) = 0.001$, $P(Y = 0.5) = P(Y = 1.5) = 0.216$ and $P(Y = 1) = 0.566$. By computing the expectations of the convex functions $|1-t|$ and $(1-t)^2$, one has that

$$E[|1-Y|] = 0.218 > 0.2 = E[|1-X_1|],$$

X_1 \ (X_2,X_3)	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$	$(2,0)$	$(0,2)$	$(2,1)$	$(1,2)$	$(2,2)$
0	0.001	0.008	0.008	0.064	0.001	0.001	0.008	0.008	0.001
1	0.008	0.064	0.064	0.518	0.008	0.008	0.064	0.064	0.008
2	0.001	0.008	0.008	0.064	0.001	0.001	0.008	0.008	0.001

Table 5.2: Probabilities associated with the possible values of the random vector (X_1,X_2,X_3) [41].

and

$$E\left[(1-Y)^2\right] = 0.11 < 0.2 = E\left[(1-X_1)^2\right].$$

Therefore, neither $\frac{1}{2}X_{(1)} + \frac{1}{2}X_{(3)} \leq_{cx} X_1$ *nor* $\frac{1}{2}X_{(1)} + \frac{1}{2}X_{(3)} \geq_{cx} X_1$ *are satisfied.*

In addition, it is clear that Theorem 5.6 holds for the important family of trimmed means and, in particular, for the median.

Corollary 5.9 *Let* $\vec{X}$ *be a random vector with symmetric independent and identically distributed components. Then, any trimmed mean is smaller in convex order than the components of* $\vec{X}$.

Corollary 5.10 *Let* $\vec{X}$ *be a random vector with symmetric independent and identically distributed components. Then, the sample median is smaller in convex order than the components of* $\vec{X}$.

5.1.3 Reduction of variability of idempotent nullnorms

A nullnorm N_a has an annihilator element $a \in [0,1]$ such that, if $a \in \{x_1,\ldots,x_n\}$, then $N_a(x_1,\ldots,x_n) = a$. In this sense, and keeping in mind the left side of Figure 2.1, one can expect, for a random vector of a big dimension, a high probability of having a as the output value. In this direction, given a sequence $(X_n, n \in \mathbb{N})$ of independent and identically distributed random variables, consider the sequence $(N_a(X_1,\ldots,X_n), n \in \mathbb{N})$. A question related to the variability of $N_a(X_1,\ldots,X_n)$ is to compute the probability of having an update, that is

$$P(N_a(X_1,\ldots,X_n) \neq N_a(X_1,\ldots,X_{n-1})).$$

Theorem 5.11 *[30] Let $(X_n, n \in \mathbb{N})$ be a sequence of independent random variables sharing the same continuous distribution on $[0,1]$ and let $N_a : \cup_{n\in\mathbb{N}}[0,1]^n \to [0,1]$ be an idempotent nullnorm with annihilator element $a \in [0,1]$. Then,*

$$P\left(N_a(X_1,\ldots,X_n) \neq N_a(X_1,\ldots,X_{n-1})\right) =$$
$$= p_a^{n-1}\left(\frac{p_a}{n} + (1-p_a)\right) + (1-p_a)^{n-1}\left(\frac{1-p_a}{n} + p_a\right),$$

for any $n \in \mathbb{N}$ with $n \geq 2$ where $p_a = P(X_1 \leq a)$.

Proof: $(N_a(X_1,\ldots,X_n), n \in \mathbb{N})$ is a discrete-time continuous-state Markov chain with absorbent state a by applying Propositions 3.44 and 3.45. Then, $N_a(X_1,\ldots,X_n) \neq N_a(X_1,\ldots,X_{n-1})$ can only occur if $X_1,\ldots,X_{n-1} < a$ or $X_1,\ldots,X_{n-1} > a$.

For the first case, that has probability p_a^{n-1}, if $X_n < a$, then $A^n(X_1,\ldots,X_n) = \min(X_1,\ldots,X_n)$ and $P\left(N_a(X_1,\ldots,X_n) \neq N_a(X_1,\ldots,X_{n-1})\right) = \frac{1}{n}$. If $X_n > a$, then it holds $P\left(N_a(X_1,\ldots,X_n) \neq N_a(X_1,\ldots,X_{n-1})\right) = 1$. Computing the whole probability, the result is $p_a^{n-1}\left(\frac{p_a}{n} + (1-p_a)\right)$. Similarly, for the second case, one has $(1-p_a)^{n-1}\left(\frac{1-p_a}{n} + p_a\right)$. The result holds by summing both probabilities. ∎

The last result implies that the sequence $(N_a(X_1,\ldots,X_n), n \in \mathbb{N})$ converges almost surely to a. Therefore, nullnorms, at least asymptotically, reduce dispersion for independent and identically distributed random variables.

5.1.3.1 Symmetric case and convex order

In the following, relax the independence assumption for $(X_n, n \in \mathbb{N})$, but consider still that the random variables have the same distribution. Similarly to the case of OWA operators, an additional symmetry condition is necessary to preserve the same expectation when the nullnorm is applied.

Theorem 5.12 *[41] Let N_a be an idempotent nullnorm with annihilator element $a \in [0,1]$. Let $(X_i, i \in \mathbb{N})$ be a sequence of random variables taking values on $[0,1]$ such that:*

1. *All random variables have the same distribution,*

2. *All random variables are symmetric with respect to a,*

3. *For any $n \in \mathbb{R}$, the copula C_n and survival copula $\bar{C}_n$ of $(X_1,\ldots,X_n)$ satisfy $C_n(t,\ldots,t) = \bar{C}_n(t,\ldots,t)$ for all $t \in [0,1]$.*

Then,

$$N_a(X_1,\ldots,X_n) \leq_{cx} N_a(X_1,\ldots,X_{n-1}) \leq_{cx} X_j,$$

for any $n, j \in \mathbb{N}$ such that $j \leq n$.

Proof: Denote as F_n and $\bar{F}_n$ the multivariate distribution function and the survival function of $(X_1,\ldots,X_n)$. Since all the components of $\vec{X}$ have the same distribution function F, one can write $F_n(t_1,\ldots,t_n) = C_n(F(t_1),\ldots,F(t_n))$ and $\bar{F}_1(t_1,\ldots,t_n) = \bar{C}(1-F(t_1),\ldots,1-F(t_n))$. In addition, denote as $F_{a,n}$ the distribution function of $N_a(X_1,\ldots,X_n)$.

If $N_a(X_1,\ldots,X_n) < a$, then it holds that $X_1,\ldots,X_n < a$ and $N_a(X_1,\ldots,X_n) = \max(X_1,\ldots,X_n)$. Then, $F_{a,n}(t) = F_n(t,\ldots,t) = C_n(F(t),\ldots,F(t))$ for $t < a$. Similarly, if one has that $N_a(X_1,\ldots,X_n) \geq a$, then $X_1,\ldots,X_n \geq a$ and $N_a(X_1,\ldots,X_n) = \min(X_1,\ldots,X_n)$. Then, $F_{a,n}(t) = 1-\bar{F}_n(t,\ldots,t) = 1-\bar{C}_n(1-F(t),\ldots,1-F(t))$ for $t \geq a$.

Since $F(a) = 0.5$, $F_{a,n}$ can be expressed as $F_{a,n}(t) = h_n(F(t))$ with $h_n : [0,1] \to [0,1]$ being

$$h_n(t) = \begin{cases} C_n(t,\ldots,t) & \text{if } t < 0.5, \\ 1-\bar{C}_n(1-t,\ldots,1-t) & \text{if } t \geq 0.5. \end{cases}$$

Consider a second distortion h'_n defined as

$$h'_n(t) = \begin{cases} C_n(t,\ldots,t) & \text{if } t < 0.5, \\ 0.5 & \text{if } t = 0.5 \\ 1-\bar{C}_n(1-t,\ldots,1-t) & \text{if } t > 0.5. \end{cases}$$

Since $C_n(t,\ldots,t) \leq C_n(t,\ldots,t,1) = C_{n-1}(t,\ldots,t)$ for any $t \in [0,1]$, where C_{n-1} is the copula of $(X_1,\ldots,X_{n-1})$, then one has that $h'_n(t) \leq h'_{n-1}(t)$ for any $t \in [0,1]$. In addition, by using the assumption that $C_n(t,\ldots,t) = \bar{C}_n(t,\ldots,t)$ for any $t \in [0,1]$ and

$n \in \mathbb{N}$, h_n' is centrally symmetric. Moreover, h_n' and h_n are equal almost everywhere for any $n \in \mathbb{N}$. Then, applying Theorem 5.1, one has the inequality

$$\int_{-\infty}^{x} F_{a,n}(t)dt = \int_{-\infty}^{x} h_n(F(t))dt = \int_{-\infty}^{x} h_n'(F(t))dt$$
$$\leq \int_{-\infty}^{x} h_{n-1}'(F(t))dt = \int_{-\infty}^{x} h_{n-1}(F(t))dt = \int_{-\infty}^{x} F_{a,n-1}(t)dt,$$

for any $x \in \mathbb{R}$.

As a consequence of symmetry, $E[N(X_1,\ldots,X_n)] = E[N(X_1,\ldots,X_{n-1})] = a$. Then, apply Theorem 2.117 to conclude that $N_a(X_1,\ldots,X_n) \leq_{cx} N_a(X_1,\ldots,X_{n-1})$.

In addition, the inequality $N_a(X_1,\ldots,X_n) \leq_{cx} X_1$ follows by the associativity of the nullnorm. Then, since the random variables in the sequence $(X_n, n \in \mathbb{N})$ have the same distribution, $N_a(X_1,\ldots,X_n) \leq_{cx} X_j$ for any $j \in [n]$. ∎

The last result implies that the random sequence $(N_a(X_1,\ldots,X_n), n \in \mathbb{N})$ is decreasing with respect to the convex order. Moreover, it holds that $N_a(X_1,\ldots,X_n) \leq_{cx} X_1$ for any $n \in \mathbb{N}$. The condition on the copula may seem too restrictive. However, there are many examples of families of copulas fulfilling this property.

Example 5.13 *[41] For a copula C_n and its corresponding survival copula $\bar{C}_n$, a sufficient condition to fulfill the property $C_n(t,\ldots,t) = \bar{C}_n(t,\ldots,t)$ for any $t \in [0,1]$ is to satisfy $C_n(x_1,\ldots,x_n) = \bar{C}_n(x_1,\ldots,x_n)$ for any $x_1,\ldots,x_n \in [0,1]$, which is equivalent to have the density copula (when exists) fulfilling $c_n(x_1,\ldots,x_n) = c_n(1-x_1,\ldots,1-x_n)$ for any $x_1,\ldots,x_n \in [0,1]$. In particular,*

- *For the independent copula $c_n(x_1,\ldots,x_n) = 1$, so it is straightforward that $c_n(x_1,\ldots,x_n) = c(1-x_1,\ldots,1-x_n)$,*
- *For the Gaussian copula, see Example 2.68, since the standard Gaussian distribution is symmetric with respect to 0, one has that $\Phi^{-1}(1-t) = -\Phi^{-1}(t)$ and therefore $c(x_1,\ldots,x_n) = c(1-x_1,\ldots,1-x_n)$,*
- *Similarly, for the T-copula, see again Example 2.68, using the symmetry of the standard Student's t-distribution with respect to 0 and the symmetry of any multivariate Student's t-distribution with respect to its mean vector, it can be verified that $c_n(x_1,\ldots,x_n) = c_n(1-x_1,\ldots,1-x_n)$.*

As mentioned in the previous example, the condition $C_n(t,\ldots,t) = \bar{C}_n(t,\ldots,t)$ for any $t \in [0,1]$ is fulfilled when $C_n(x_1,\ldots,x_n) = \bar{C}_n(x_1,\ldots,x_n)$ for any $x_1,\ldots,x_n \in [0,1]$, which is known as radial symmetry in Copula Theory (see page 32 in [118]). However, there are cases in which the first condition holds without the necessity of the second.

Example 5.14 *[41] Consider the copula C with density c given by*

$$c(x,y) = \begin{cases} \frac{4}{3} & \text{if } x \leq \frac{3}{4} \text{ and } y \geq \frac{1}{4}, \\ 4 & \text{if } x \geq \frac{3}{4} \text{ and } y \leq \frac{1}{4}. \end{cases}$$

Then, the expression of the copula and the survival copula are

$$C(x,y) = \begin{cases} 0 & \text{if } x \leq \frac{3}{4} \text{ and } y \leq \frac{1}{4}, \\ 4\left(x-\frac{3}{4}\right)y & \text{if } x \geq \frac{3}{4} \text{ and } y \leq \frac{1}{4}, \\ \frac{4}{3}x\left(y-\frac{1}{4}\right) & \text{if } x \leq \frac{3}{4} \text{ and } y \geq \frac{1}{4}, \\ 4\left(x-\frac{3}{4}\right)y+\frac{4}{3}x\left(y-\frac{1}{4}\right)+ & \text{if } x \geq \frac{3}{4} \text{ and } y \geq \frac{1}{4}, \end{cases}$$

and

$$\bar{C}(x,y) = \begin{cases} 0 & \text{if } x \leq \frac{1}{4} \text{ and } y \leq \frac{3}{4}, \\ \frac{4}{3}\left(x-\frac{1}{4}\right)y & \text{if } x \geq \frac{1}{4} \text{ and } y \leq \frac{3}{4}, \\ 4x\left(y-\frac{3}{4}\right) & \text{if } x \leq \frac{1}{4} \text{ and } y \geq \frac{3}{4}, \\ \frac{4}{3}\left(x-\frac{1}{4}\right)y+4x\left(y-\frac{3}{4}\right)+ & \text{if } x \geq \frac{1}{4} \text{ and } y \geq \frac{3}{4}. \end{cases}$$

It is clear that $C(x,y) \neq \bar{C}(x,y)$ for all $x,y \in [0,1]$ since, for instance, $C(\frac{1}{3},\frac{1}{2}) = \frac{1}{9}$ and $\bar{C}(\frac{1}{3},\frac{1}{2}) = \frac{1}{18}$. On the other hand, on the diagonal they assume the same values,

$$C(t,t) = \bar{C}(t,t) = \begin{cases} 0 & \text{if } t \leq \frac{1}{4}, \\ \frac{4}{3}t\left(t-\frac{1}{4}\right) & \text{if } \frac{1}{4} < t \leq \frac{3}{4}, \\ \frac{4}{3}\left(t-\frac{1}{4}\right)t+4t\left(t-\frac{3}{4}\right) & \text{if } \frac{3}{4} < t. \end{cases}$$

5.1.3.2 Non-symmetric case and increasing convex order

Note that the symmetric distributions for the components of $\vec{X}$ are a restrictive assumption in Theorem 5.12 for many applicative purposes, as well as the assumption that the annihilator element a must be the median. This section is devoted to

the case in which the distribution is non-symmetric, showing that in this case the increasing convex order can be achieved under suitable assumptions. The following preliminary result will be used later.

Lemma 5.15 *[41] Let N_a be an idempotent nullnorm with annihilator element $a \in [0,1]$. Let $(X_i, i \in \mathbb{N})$ be a sequence of random variables with the same distribution taking values in $[0,1]$. Then,*

$$N_a(X_1,\ldots,X_n) \leq_{icx} N_a(X_1,\ldots,X_{n-1}) \Longleftrightarrow E[N_a(X_1,\ldots,X_n)] \leq E[N_a(X_1,\ldots,X_{n-1})].$$

Proof: Note that the implication $\Rightarrow$ holds by the definition of the increasing convex order. For the second one, recall that the survival function of $N_a(X_1,\ldots,X_n)$, denoted as $\bar{F}_{a,n}$, is

$$\bar{F}_{a,n}(t) = \begin{cases} 1 - C_n(F(t),\ldots,F(t)) & \text{if } t < a, \\ \bar{C}_n(1-F(t),\ldots,1-F(t)) & \text{if } t \geq a, \end{cases}$$

where F denotes the distribution function of X_1.

Notice that, by the monotonicity of C and $\bar{C}$, $C_n(t,\ldots,t) \leq C_n(t,\ldots,t,1) = C_{n-1}(t,\ldots,t)$ and $\bar{C}_n(t,\ldots,t) \leq \bar{C}_n(t,\ldots,t,1) = \bar{C}_{n-1}(t,\ldots,t)$ for any $t \in [0,1]$.

Then, $\bar{F}_{a,n}(t) \geq \bar{F}_{a,n-1}(t)$ for $t < a$ and $\bar{F}_{a,n}(t) \leq \bar{F}_{a,n-1}(t)$ for $t \geq a$. Therefore, for $x \geq a$ then one has

$$\int_x^1 \bar{F}_{a,n}(t)dt \leq \int_x^1 \bar{F}_{a,n-1}(t)dt.$$

If $x \leq a$, notice that $E[N_a(X_1,\ldots,X_n)] \leq E[N_a(X_1,\ldots,X_{n-1})]$ by assumption, which implies that $\int_0^1 \bar{F}_{a,n}(t)dt \leq \int_0^1 \bar{F}_{a,n-1}(t)dt$. Moreover, since $\int_0^x \bar{F}_{a,n}(t)dt \geq \int_0^x \bar{F}_{a,n-1}(t)dt$, then

$$\begin{aligned} \int_x^1 \bar{F}_{a,n}(t)dt &= \int_0^1 \bar{F}_{a,n}(t)dt - \int_0^x \bar{F}_{a,n}(t)dt \leq \\ \leq \int_0^1 \bar{F}_{a,n-1}(t)dt &- \int_0^x \bar{F}_{a,n-1}(t)dt = \int_x^1 \bar{F}_{a,n-1}(t)dt. \end{aligned}$$

Finally, the result holds as a consequence of Theorem 2.109. ∎

The previous property affirms that the comparison in the increasing convex order between the nullnorms with n and $n-1$ inputs is equivalent to the comparison among the corresponding expectations. In general, the latter comparison is not satisfied for any possible annihilator element a, however, there always exists an upper bound for a such that the inequality is satisfied, as shown in the next result.

Theorem 5.16 *[41] Let N_a be the idempotent nullnorm with annihilator element $a \in [0,1]$. Let X_1 and X_2 be random variables defined in $[0,1]$. Then, there always exists $a' \in [0,1]$ such that $N_a(X_1,X_2) \leq_{icx} X_1$ for any $a \in [0,1]$ satisfying $a \leq a'$.*

Proof: Denote the distribution functions of X_1 and X_2 as F_1 and F_2 and the copula and the survival copula of (X_1,X_2) as C and $\bar{C}$. Then, the survival function of $N_a(X_1,X_2)$, denoted as $\bar{F}_a$, equals

$$\bar{F}_a(t) = \begin{cases} 1 - C(F_1(t),F_2(t)) & \text{if } t < a, \\ \bar{C}(\bar{F}_1(t),\bar{F}_2(t)) & \text{if } t \geq a. \end{cases} \tag{5.3}$$

Therefore, the expectation of $N_a(X_1,X_2)$ can the written as a function of a as

$$E[N_a(X_1,X_2)] = \int_0^a (1 - C(F_1(t),F_2(t)))\,dt + \int_a^1 \left(\bar{C}(\bar{F}_1(t),\bar{F}_2(t))\right) dt.$$

Trivially, $E[N_0(X_1,X_2)] = E[\min(X_1,X_2)]$ and $E[N_1(X_1,X_2)] = E[\max(X_1,X_2)]$. Moreover, such a function of a is continuous with derivative

$$\frac{d}{da}E[N_a(X_1,X_2)] = \frac{d}{da}\int_0^a (1 - C(F_1(t),F_2(t)))\,dt +$$
$$+\frac{d}{da}\int_a^1 \left(\bar{C}(1-F_1(t),1-F_2(t))\right) dt = 1 - C(F_1(a),F_2(a)) - \bar{C}(\bar{F}_1(a),\bar{F}_2(a)).$$

Since $C(F_1(a),F_2(a)) + \bar{C}(\bar{F}_1(a),\bar{F}_2(a)) = P(X_1 \leq a, X_2 \leq a) + P(X_1 > a, X_2 > a) \leq 1$ for any $a \in [0,1]$, the latter derivative is positive, thus $E[N_a(X_1,X_2)]$ is increasing as a function of a.

Moreover, since $E[N_0(X_1,X_2)] = E[\min(X_1,X_2)] \leq E[X_1] \leq E[\max(X_1,X_2)] = E[N_1(X_1,X_2)]$ and $E[N_a(X_1,X_2)]$ is continuous as a function of a, there exists $a' \in [0,1]$ such that $E[N_{a'}(X_1,X_2)] = E[X_1]$ and $E[N_a(X_1,X_2)] \leq E[X_1]$ for any $a \leq a'$. The result holds by Lemma 5.15. ∎

Notice that, by using the associative property, the latter result can be extended for any number of possible inputs as follows.

Corollary 5.17 *[41] Let N_a be the idempotent nullnorm with annihilator element $a \in [0,1]$. Let $X_1,\ldots,X_n$ be random variables defined in $[0,1]$. Then, there always exists $a' \in [0,1]$ such that $N_a(X_1,\ldots,X_n) \leq_{icx} N_a(X_1,\ldots,X_{n-1})$ for any $a \in [0,1]$ satisfying $a \leq a'$.*

Proof: Notice that $N_a(X_1,\ldots,X_{n-1},X_n) = N_a(N_a(X_1,\ldots,X_{n-1}),X_n)$. Then, apply Theorem 5.16. ∎

One of the main limitations of the latter result is that it does not give a value for a'. In the next result, it is shown that, when the inputs are independent, have the same distribution and under some additional conditions, the value of a' is always greater than the median of the components of $\vec{X}$. Thus, $N_a(X_1,\ldots,X_n) \leq_{icx} N_a(X_1,\ldots,X_{n-1})$ holds for any a smaller than the median of X_1.

Theorem 5.18 *[41] Let N_a be the idempotent nullnorm with annihilator element $a \in [0,1]$. Let $\vec{X} = (X_1,\ldots,X_n)$ be a random vector with independent and identically distributed components having support $[0,1]$. If the survival function $\bar{F}$ of the components is convex on $[0,1]$ then*

$$N_a(X_1,\ldots,X_n) \leq_{icx} X_j, \forall j \in [n],$$

for any $a \in [0,1]$ such that $a \leq m$, where m denotes the median of X_1.

Proof: As a consequence of Lemma 5.15, the statement holds if $E[N_m(X_1,\ldots,X_n)] \leq E[X_j]$. Notice that, if $\bar{F}$ is convex on $[0,1]$, then it is continuous and the corresponding density function f is decreasing on $[0,1]$. Since independence between the random variables is assumed, Equation (5.3) reads as

$$\bar{F}_m(t) = \begin{cases} 1 - F(t)^n & \text{if } t < m, \\ \bar{F}(t)^n & \text{if } t \geq m, \end{cases}$$

and, therefore, the expectations are ordered if and only if

$$E[N_m(X_1,\ldots,X_n)] = \int_0^m (1-F(t)^n)\,dt + \int_m^1 \bar{F}(t)^n dt \leq$$
$$\leq \int_0^m (1-F(t))\,dt + \int_m^1 \bar{F}(t)dt = E[X_1].$$

The latter inequality can be rewritten as

$$\int_0^m (1-F(t)^n - 1 - F(t))\,dt \leq \int_m^1 (\bar{F}(t) - \bar{F}(t)^n)\,dt,$$

which in turn is equivalent to

$$\int_0^m F(t)\left(1-F(t)^{n-1}\right)dt \leq \int_m^1 \bar{F}(t)\left(1-\bar{F}(t)^{n-1}\right)dt.$$

Since f is decreasing, the median m is smaller than 0.5. Computing the derivative of $F(m-t)\left(1-F(m-t)^{n-1}\right)$ and of $\bar{F}(m+t)\left(1-\bar{F}(m+t)^{n-1}\right)$ for $t \in [0,m]$, one gets

$$\frac{d}{dt}\left[F(m-t)\left(1-F(m-t)^{n-1}\right)\right] = -f(m-t)\left(1-nF(m-t)^{n-1}\right),$$
$$\frac{d}{dt}\left[\bar{F}(m+t)\left(1-\bar{F}(m+t)^{n-1}\right)\right] = -f(m+t)\left(1-n\bar{F}(m+t)^{n-1}\right).$$

Notice that, since f is decreasing, then $f(m-t) \geq f(m+t)$ for any $t \in [0,m]$. In addition,

$$F(m-t) = P(X \leq m) - P(X \in (m-t,m]) = 0.5 - P(X \in (m-t,m]),$$
$$\bar{F}(m+t) = P(X > m) - P(X \in (m,m+t]) = 0.5 - P(X \in (m,m+t]).$$

Since f is decreasing, it follows that $P(X \in (m-t,m]) \geq P(X \in (m,m+t])$, thus $F(m-t) \leq \bar{F}(m+t)$ and $\left(1-nF(m-t)^{n-1}\right) \geq \left(1-n\bar{F}(m+t)^{n-1}\right)$ for any $t \in [0,m]$. Therefore, it is concluded that

$$\frac{d}{dt}\left[F(m-t)\left(1-F(m-t)^{n-1}\right)\right] \leq \frac{d}{dt}\left[\bar{F}(m+t)\left(1-\bar{F}(m+t)^{n-1}\right)\right],$$

for any $t \in [0,m]$. Moreover, using that $F(m) = \bar{F}(m) = 0.5$ it follows $F(m-t)\left(1-F(m-t)^{n-1}\right) \leq \bar{F}(m+t)\left(1-\bar{F}(m+t)^{n-1}\right)$ for any $t \in [0,m]$. Thus, since $2m \leq 1$,

$$\int_0^m F(t)\left(1-F(t)^{n-1}\right)dt \leq \int_m^{2m} \bar{F}(t)\left(1-\bar{F}(t)^{n-1}\right)dt \leq \int_m^1 \bar{F}(t)\left(1-\bar{F}(t)^{n-1}\right)dt.$$

Finally, since $E[N_a(X_1,\ldots,X_n)]$ increases in a, applying Lemma 5.15 it holds $N_a(X_1,\ldots,X_n) \leq_{icx} N_m(X_1,\ldots,X_n) \leq_{icx} X_j$ for any $j \in [n]$. ∎

5.1.4 The case of uninorms

The structure of uninorms is more complex than the structure of nullnorms, mainly due to the choice of the function g in Proposition 2.21. If one tries to replicate Theorem 5.11, it is difficult to obtain a simple formula for the probability of having two different subsequent values. However, some bounds can be given.

Theorem 5.19 *[30] Let $(X_n, n \in \mathbb{N})$ be a sequence of independent random variables sharing the same continuous distribution on $[0,1]$ and $U_e : [0,1]^2 \to [0,1]$ an idempotent uninorm with neutral element $e \in [0,1]$. Then,*

$$\frac{p_e^n + (1-p_e)^n}{n} \leq P\left(U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})\right) \leq 1 - \frac{p_e^n + (1-p_e)^n}{n},$$

for any $n \in \mathbb{N}$ with $n \geq 2$ where $p_e = P(X_1 \leq e)$.

Proof: Using Proposition 2.21, consider the sets of points in the unit square associated with the values of $U_e(X_1,\ldots,X_{n-1})$ and X_n for which $U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})$. That is, $\{x,y \in [0,1] \mid x > y, y < g(x)\}$ and $\{x,y \in [0,1] \mid x < y, y > g(x)\}$. The remaining points associated with $U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})$ have Lebesgue measure 0, thus they are negligible when considering independent continuous random variables.

For the lower bound, if one wants to minimize the sets of points for which $U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})$, g in Proposition 2.21 can be chosen to be $g(x) = 1$ if $x \in [0,e)$, $g(e) = e$ and $g(x) = 0$ if $x \in (e,1]$. Suppose $U_e(X_1,\ldots,X_{n-1}) < e$. Since U_e is idempotent, therefore internal, the probability of $U_e(X_1,\ldots,X_{n-1}) < e$ is greater than the probability of the maximum, which is p_e^{n-1}. In that case, $U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})$ if and only if $X_n < U_e(X_1,\ldots,X_{n-1})$. Firstly, it is necessary $X_n < e$, which has probability p_e and then having, given that $U_e(X_1,\ldots,X_{n-1}) < e$ and $X_n < e$, that $X_n < U_e(X_1,\ldots,X_{n-1})$. This probability is greater than the one of $X_n < \min(X_1,\ldots,X_{n-1})$, which is $\frac{1}{n}$. Multiplying the terms, the resulting

probability is $\frac{p_e^n}{n}$. For the case $U_e(X_1,\ldots,X_{n-1}) > e$, one similarly obtains $\frac{(1-p_e)^n}{n}$. The lower bound is reached by summing both terms.

For the upper bound, if one wants to maximize the sets of points for which $U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})$, g can be chosen to be $g(x) = e$ for all $x \in [0,1]$. Suppose $X_n < e$, which happens with probability p_e. Therefore, $U_e(X_1,\ldots,X_n) \neq U_e(X_1,\ldots,X_{n-1})$ if and only if $X_n < U_e(X_1,\ldots,X_{n-1})$. That always happens if $U_e(X_1,\ldots,X_{n-1}) > e$, for which the probability is upper bounded by $P(\max(X_1,\ldots,X_{n-1}) > e) = 1 - p_e^{n-1}$. For the other case, if $U_e(X_1,\ldots,X_{n-1}) < e$, an upper bound of the probability can be computed considering the one of $X_n < U_e(X_1,\ldots,X_{n-1})$ when $X_n < e$ and $\max(X_1,\ldots,X_n) < e$, which is $1 - \frac{1}{n}$. Summarizing all the probabilities, it is obtained $p_e\left(1 - p_e^{n-1} + p_e^{n-1}\left(1 - \frac{1}{n}\right)\right) = p_e - p_e^n\frac{1}{n}$. For the case $X_n > e$, one similarly obtains $1 - p_e - (1 - p_e^n)\frac{1}{n}$. The upper bound is reached by summing both terms. ∎

Admittedly, the bounds given in the latter result are not tight, mainly because the choices of g in the proof lead to upper and lower bounds for the probabilities, but they are not associated with any uninorm, because of the lack of symmetry. Moreover, asymptotically the bounds tend to 0 and 1, so it is not possible to obtain any information about a possible reduction of the variability. However, other techniques can be used.

One can think of uninorms as functions that return the maximum when the variables take big values and the minimum when the values are smaller (with some intermediate cases). In this sense, a convergence to 1 (or 0), if all variables take enough big (or small) values, can be expected.

Proposition 5.20 *[41] Let U be a uninorm with neutral element e and $(X_n, n \in \mathbb{N})$ a sequence of independent and identically distributed random variables such that $P(X_1 > e) = 1$. Then, $U(X_1,\ldots,X_n) \to_{a.s.} \sup S(X_1)$.*

Proof: Notice that, if $X_1,\ldots,X_n > e$, then $U(X_1,\ldots,X_n) = \max(X_1,\ldots,X_n)$. Therefore, it is clear that $U(X_1,\ldots,X_n) \to_{a.s.} \sup S(X_1)$. ∎

A similar result, with an equivalent proof, can be stated for random variables assuming small values.

Proposition 5.21 *[41] Let U be a uninorm with neutral element e and $(X_n, n \in \mathbb{N})$ a sequence of independent and identically distributed random variables such that $P(X_1 < e) = 1$. Then, $U(X_1, \ldots, X_n) \to_{a.s.} \inf S(X_1)$.*

In both cases, although they are not inequalities in terms of the convex order, there is, at least asymptotically, a reduction of the variability, since the limit is a degenerate random variable.

However, this is not the case in all situations. If the distribution is evenly distributed below and above the function g, some part of the distribution will increase, while the other part will decrease, leading to an increase in variability. Although a general result for this behavior cannot be proven, it is possible to consider particular uninorms to illustrate this property.

Proposition 5.22 *[41] Let X_1 and X_2 be two independent, continuous, symmetric with respect to 0.5 and identically distributed random variables. Let U be the uninorm defined by considering $g(x) = 1 - x$ in Proposition 2.21. Then,*

$$U(X_1, X_2) \geq_{cx} X_1.$$

Proof: Notice that, since the random variables are continuous and the set

$$\{(x_1, x_2) \in [0,1]^2 \mid x_1 = 1 - x_2\},$$

has Lebesgue measure 0, the behavior of the uninorm over it is negligible. Compute the density function f_U of $U(X_1, X_2)$ using the density function and the distribution function of X_1, denoted by f and F. Using Proposition 2.21, the preimage of U is

$$U(x_1, x_2) = y \iff \begin{cases} y = x_1 < x_2 < 1 - y \text{ or } y = x_2 < x_1 < 1 - y & \text{if } y < 0.5, \\ y = x_1 > x_2 > 1 - y \text{ or } y = x_2 > x_1 > 1 - y & \text{if } y > 0.5. \end{cases}$$

Therefore,

$$f_U(y) = \begin{cases} 2\int_y^{1-y} f(y) f(t) dt = f(t)\left(F(1-y) - F(y)\right) & \text{if } y < 0.5, \\ 2\int_{1-y}^{y} f(y) f(t) dt = f(t)\left(F(y) - F(1-y)\right) & \text{if } y > 0.5. \end{cases}$$

Using the symmetry of the random variables, it holds that $F(1-y)-F(y) = 1-2F(y)$ and, therefore, $f_U(y) = f(y)|1-2F(y)|$. The associated distribution function is

$$F_U(y) = \begin{cases} \frac{1}{2} - \frac{1}{2}(1-2F(y))^2 & \text{if } y < 0.5, \\ \frac{1}{2} + \frac{1}{2}(1-2F(y))^2 & \text{if } y \geq 0.5. \end{cases}$$

Then, $F_U(y) = h(F(y))$ with h being the distortion given by:

$$h(t) = \begin{cases} \frac{1}{2} - \frac{1}{2}(1-2t)^2 & \text{if } t < 0.5, \\ \frac{1}{2} + \frac{1}{2}(1-2t)^2 & \text{if } t \geq 0.5. \end{cases}$$

Notice that h is centrally symmetric and fulfills $h(t) \geq t$ for any $t \in [0, 0.5]$. Then, applying Lemma 5.1, it is concluded that $U(X_1, X_2) \geq_{cx} X_1$. ∎

Unfortunately, in many cases, one cannot say anything about the relationship between the inputs and the output with respect to the convex order.

Example 5.23 *[41] Let X_1 and X_2 be two independent standard uniform random variables and U the uninorm given by*

$$U(x_1, x_2) = \begin{cases} \min(x_1, x_2) & \text{if } x_1, x_2 < 0.5, \\ \max(x_1, x_2) & \text{elsewhere.} \end{cases}$$

Then, the survival function of $U(X_1, X_2)$ fulfills $\bar{F}_U(t) = 0.25(1-2t)^2 + 0.75$ if $t < 0.5$ and $\bar{F}_U(t) = 1 - t^2$ if $t \geq 0.5$. Thus,

$$E[U(X_1, X_2)] = \int_0^{0.5} \left(0.25(1-2t)^2 + 0.75\right) dt + \int_{0.5}^1 \left(1 - t^2\right) dt = \frac{5}{8}.$$

Since the expectation of a standard uniform is $\frac{1}{2}$, neither $U(X_1, X_2) \leq_{cx} X_1$ nor $U(X_1, X_2) \geq_{cx} X_1$ hold.

5.2 Variability measures and variability orders

Some aggregation functions are defined as solutions of some minimization problems that involve the so-called penalty aggregation functions (see [75, 78]). In particular, the value of the aggregation is computed by minimizing the penalty function

between a real number, which will be the output value of the aggregation, and a given vector, which is the input. For instance, if the penalty function is the sum of the squares of the difference between the real number and the values in the given vector, the associated aggregation function is the arithmetic mean. In the next result, two relevant cases are shown.

Theorem 5.24 *[78] Consider a weighting vector $\vec{w} \in \mathbb{R}^n$. Then,*

(a) $\sum_{i=1}^{n} w_i x_i = \arg\min_{y\in\mathbb{R}} \sum_{i=1}^{n} w_i (x_i - y)^2$,

(b) $\sum_{i=1}^{n} w_i x_{(i)} = \arg\min_{y\in\mathbb{R}} \sum_{i=1}^{n} w_i (x_{(i)} - y)^2$.

In general, penalty functions measure the distance between a real number and the values of a vector. Moreover, when the penalty function takes its minimum value, it often equals a well-known variability measure from Statistics. Returning to the previous example, the sum of the squared distance between the values of a vector and its arithmetic mean is just the sample variance (multiplied by a constant). Therefore, from the solution of finding the minimum of a penalty function, two values appear; the minimum, which is a variability measure of the considered vector, and the point where the minimum is attained, which is the value of the aggregation function.

In this direction, it is interesting to know when the penalty function evaluated in its minimum takes greater or smaller values, since it is a measure of the similarity between the aggregated and the initial values. In this section, some stochastic inequalities between variability measures which are penalty functions evaluated in their minimum are provided. In addition, other variability measures that are not related to penalty functions, such as the Gini mean distance, are also considered.

As a first simple example, let $\vec{X}$ and $\vec{Y}$ be two random vectors with independent and identically distributed components. Assume that $\mathrm{Var}(X_1) \leq \mathrm{Var}(X_2)$. Then, it is well-known that the expectations of the sample variances are also ordered, i.e., $E[\hat{\sigma}^2(\vec{X})] \leq E[\hat{\sigma}^2(\vec{Y})]$, since the sample variance is an unbiased estimator of the variance.

However, one may wonder if they are also ordered in the usual stochastic order ($\hat{\sigma}^2(\vec{X}) \leq_{st} \hat{\sigma}^2(\vec{Y})$). Apart from obvious theoretical motivations and the application

in aggregation based on penalty functions, there are several reasons why a stochastic inequality as $\hat{\sigma}^2(\vec{X}) \leq_{st} \hat{\sigma}^2(\vec{Y})$ can be more useful than an inequality between the corresponding expected values. For instance, suppose that one wants to determine the confidence intervals for the expectation of the marginals of $\vec{X}$ and $\vec{Y}$. For the same confidence level and big enough sample size (see [278]), the widths of the intervals are ordered with respect to the usual stochastic order whenever $\hat{\sigma}^2(\vec{X}) \leq_{st} \hat{\sigma}^2(\vec{Y})$ [38]. This type of results cannot be obtained with $E[\hat{\sigma}^2(\vec{X})] \leq E[\hat{\sigma}^2(\vec{Y})]$, not even for the comparison of the expectations of the widths.

Unfortunately, condition $\mathrm{Var}(X_1) \leq \mathrm{Var}(Y_1)$ is not sufficient for the comparison $\hat{\sigma}^2(\vec{X}) \leq_{st} \hat{\sigma}^2(\vec{Y})$, not even in the case of independent and identically distributed components, as illustrated in the following example.

Example 5.25 *[38] Let X_1 and X_2 be two independent random variables with standard uniform distribution and let Y_1 and Y_2 be two independent random variables with exponential distribution with parameter $\lambda = 3$, so that $Var(X_1) = \frac{1}{12}$ and $Var(Y_1) = \frac{1}{9}$. However, it is not true that $\hat{\sigma}^2(\vec{X}) \leq_{st} \hat{\sigma}^2(\vec{Y})$. In fact, note that $P[(X_1 - X_2)^2 \leq x] = 1 - (1 - \sqrt{x})^2$, $x \in [0,1]$ and $P[(Y_1 - Y_2)^2 \leq x] = 1 - e^{-\lambda\sqrt{x}}$, $x \in [0,\infty]$. Representing both distribution functions, as it can be seen in Figure 5.1, they cross. Since $\hat{\sigma}^2(\vec{X}) = \frac{1}{4}(X_1 - X_2)^2$ and $\hat{\sigma}^2(\vec{Y}) = \frac{1}{4}(Y_1 - Y_2)^2$, it is concluded that $\hat{\sigma}^2(\vec{X}) \not\leq_{st} \hat{\sigma}^2(\vec{Y})$ even with $Var(X_1) \leq Var(Y_1)$ being true.*

What has just been shown with a counterexample also applies to other kinds of variability measures such as the Gini mean difference or the range, (see Section 2.4.1). Inequalities considering the variance, the width of the support or similar variability quantities of the initial random variables do not imply the stochastic comparison of the sample measures. Therefore, stronger conditions related to stochastic orders are needed.

5.2.1 Conditions for the usual stochastic order

The aim of this section is to give sufficient conditions, related to the dispersive order, such that the variability measures estimators mentioned above can be compared in terms of the usual stochastic order.

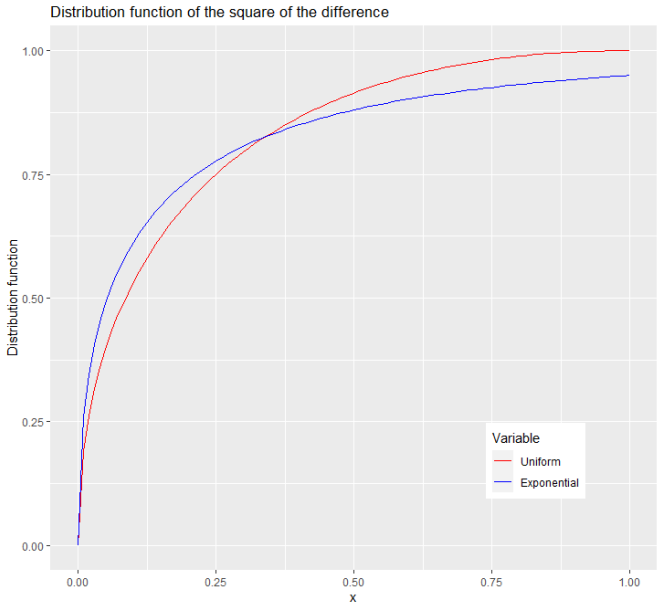

Figure 5.1: Distribution functions of the square of the difference of two independent random variables with standard uniform distribution and the square of the difference of two independent random variables with exponential distribution with parameter $\lambda = 3$ [38].

In the following, consider two random vectors in which all the components have the same distribution, within each of them. In addition, consider that both vectors have the same copula. Intuitively, they can be seen as random samples but without necessarily independent variables. Under such conditions, the dispersive order implies the existence of a contraction that links the distribution of both random vectors. In this direction, given two random variables X and Y, the following two conditions will be considered in the next statements.

(P1) Both X and Y have continuous distribution functions F and G,

(P2) The transformation $\phi = F^{-1} \circ G$ such that $X = \phi(Y)$ is strictly monotone.

Note that condition (P2) is satisfied, for example, if both X and Y have non-zero density over a convex subset of the real line, which can be different for X and Y.

Lemma 5.26 *[38] Let $\vec{X}$ be a random vector with identically distributed components and $\vec{Y}$ another random vector with identically distributed components and with the same copula as $\vec{X}$. If $X_1 \leq_{disp} Y_1$ and X_1 and Y_1 fulfill (P1) and (P2), then it holds that $\vec{X} =_{st} (\phi(Y_1), \ldots, \phi(Y_n))$ where $\phi = F^{-1} \circ G$ is a contraction and F and G are the distribution functions of X_1 and Y_1, respectively.*

Proof: Since $X_1 =_{st} \cdots =_{st} X_n$, $Y_1 =_{st} \cdots =_{st} Y_n$ and $X_1 \leq_{disp} Y_1$, then $X_i =_{st} \phi(Y_i)$ with ϕ being a contraction for any $i \in [n]$ (see the comment right after Definition 2.115). In addition, since the marginals are continuous by (P1) and ϕ is strictly increasing by (P2), the copulas of $(\phi(Y_1), \ldots, \phi(Y_n))$ and $\vec{Y}$ are the same. Therefore, $(\phi(Y_1), \ldots, \phi(Y_n))$ and $\vec{X}$ have the same copula. Moreover, since they have the same copula and the same marginals, $\vec{X} =_{st} (\phi(Y_1), \ldots, \phi(Y_n))$. ∎

Notice that the random vectors can have different marginals, the marginals inside the two random vectors are the ones which should be the same. A direct consequence of this result is that the vectors consisting of the absolute difference between the components are ordered in the usual order. Given a random vector $\vec{X}$, the notation $\vec{X}_{AD} = (|X_i - X_j|, i, j \in [n])$ will be used for the random vector consisting of the absolute difference between all the possible combinations of the components of $\vec{X}$.

Theorem 5.27 *[38] Let $\vec{X}$ be a random vector with identically distributed components and $\vec{Y}$ another random vector with identically distributed components and with the same copula as $\vec{X}$. If $X_1 \leq_{disp} Y_1$ and X_1 and Y_1 fulfill (P1) and (P2), then*

$$\vec{X}_{AD} \leq_{st} \vec{Y}_{AD}.$$

Proof: Applying Lemma 5.26, there exists an increasing contraction ϕ such that $\vec{X} =_{st} (\phi(Y_1), \ldots, \phi(Y_n))$. Then, using the definition of contraction,

$$\vec{X}_{AD} =_{st} (|\phi(Y_i) - \phi(Y_j)|, i, j \in [n]) \leq_{a.s.} \vec{Y}_{AD}.$$

The result holds by applying Theorem 2.102. ∎

In particular, as an immediate consequence of Theorem 5.27 and Proposition 2.106, the following conditions for the comparisons among sample variances and estimators of the Gini mean difference can be proved.

Corollary 5.28 *[38] Let $(X_1,\ldots,X_n)$ and $(Y_1,\ldots,Y_n)$ be two random vectors with independent and identically distributed random variables. If $X_1 \leq_{disp} Y_1$ and X_1 and Y_1 fulfill (P1) and (P2), then $\hat{\sigma}^2(\vec{X}) \leq_{st} \hat{\sigma}^2(\vec{Y})$ and $G(\vec{X}) \leq_{st} G(\vec{Y})$.*

Conditions for the stochastic ordering between weighted versions of the sample variances can also be stated. In particular, under the same conditions as the latter results, the dispersive order of the random vectors implies the usual stochastic order between these quantities.

Theorem 5.29 *[38] Let $\vec{X}$ be a random vector with identically distributed components and $\vec{Y}$ another random vector with identically distributed components and with the same copula as $\vec{X}$. Assume that X_1 and Y_1 fulfill (P1) and (P2). Then, for any weighting vector $\vec{w} \in [0,1]^n$,*

$$X_1 \leq_{disp} Y_1 \implies \sum_{i=1}^{n} w_i \left(X_i - \sum_{j=1}^{n} w_j X_j \right)^2 \leq_{st} \sum_{i=1}^{n} w_i \left(Y_i - \sum_{j=1}^{n} w_j Y_j \right)^2 .$$

Proof: Applying Lemma 5.26, there exists an increasing contraction $\phi : \mathbb{R} \to \mathbb{R}$ such that $\vec{X} =_{st} (\phi(Y_1),\ldots,\phi(Y_n))$. Define, in the same probability space where $\vec{Y}$ is defined, the vector $\widehat{\vec{X}} = (\phi(Y_1),\ldots,\phi(Y_n))$. Then,

$$\sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} w_j \widehat{X}_j \right)^2 =_{a.s.} \sum_{i=1}^{n} w_i \left(\phi(Y_i) - \sum_{j=1}^{n} w_j \phi(Y_j) \right)^2 \leq_{a.s.}$$

$$\leq_{a.s.} \sum_{i=1}^{n} w_i \left(\phi(Y_i) - \phi\left(\sum_{j=1}^{n} w_j Y_j \right) \right)^2 \leq_{a.s.} \sum_{i=1}^{n} w_i \left(Y_i - \sum_{j=1}^{n} w_j Y_j \right)^2 ,$$

where the second inequality follows from (a) in Theorem 5.24 and the third by the fact that ϕ is a contraction. Observing that

$$\sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} w_j \widehat{X}_j \right)^2 =_{st} \sum_{i=1}^{n} w_i \left(X_i - \sum_{j=1}^{n} w_j X_j \right)^2 ,$$

the result holds by applying Theorem 2.102. ∎

It must be observed that the conditions for the usual stochastic order between sample variances stated in Corollary 5.28 also follow from Theorem 5.29. Recall

that, as mentioned in Section 2.3.2.5, estimators based on order statistics are relevant in many areas. A result similar to Theorem 5.29 can also be stated when working with ordered samples.

Theorem 5.30 *[38] Let $\vec{X}$ be a random vector with identically distributed components and $\vec{Y}$ another random vector with identically distributed components and with the same copula as $\vec{X}$. Assume that X_1 and Y_1 fulfill (P1) and (P2). Then, for any weighting vector $\vec{w} \in [0,1]^n$,*

$$X_1 \leq_{disp} Y_1 \implies \sum_{i=1}^{n} w_i \left(X_{(i)} - \sum_{j=1}^{n} w_j X_{(j)} \right)^2 \leq_{st} \sum_{i=1}^{n} w_i \left(Y_{(i)} - \sum_{j=1}^{n} w_j Y_{(j)} \right)^2 .$$

Proof: Applying Lemma 5.26, $\vec{X} =_{st} (\phi(Y_1), \ldots, \phi(Y_n))$ with ϕ an increasing contraction. Since ϕ is increasing, the order statistics of $(\phi(Y_1), \ldots, \phi(Y_n))$ equal $\left(\phi(Y_{(1)}), \ldots, \phi(Y_{(n)})\right)$. Thus, $\left(X_{(1)}, \ldots, X_{(n)}\right) =_{st} \left(\phi(Y_{(1)}), \ldots, \phi(Y_{(n)})\right)$. Then, use (b) in Theorem 5.24 and proceed exactly as in Theorem 5.29. ∎

The next statements generalize Theorem 5.29 and Theorem 5.30 relaxing the condition of the same copula for $\vec{X}$ and $\vec{Y}$. In particular, the following results can be applied to random vectors that have distributions that depend on a set of random parameters, such as, for example, vectors of lifetimes described by multivariate frailty models [231] (see Example 5.33).

Theorem 5.31 *[38] Let $\vec{X}(\Theta_X)$ and $\vec{Y}(\Theta_Y)$ be two random vectors depending on two random parameters Θ_X and Θ_Y with the same support $\mathscr{T}$ such that:*

(1) Each one of $[\vec{X}(\Theta_X) \mid \Theta_X = \theta]$ and $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta]$ is a random vector with identically distributed marginals for any $\theta \in \mathscr{T}$,

(2) $[\vec{X}(\Theta_X) \mid \Theta_X = \theta_1]$, $[\vec{X}(\Theta_X) \mid \Theta_X = \theta_2]$, $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta_1]$ and $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta_2]$ have the same copula for any $\theta_1, \theta_2 \in \mathscr{T}$,

(3) $[X_i(\Theta_X) \mid \Theta_X = \theta_1] \leq_{disp} [X_i(\Theta_X) \mid \Theta_X = \theta_2]$, or $[Y_i(\Theta_Y) \mid \Theta_Y = \theta_1] \leq_{disp} [Y_i(\Theta_Y) \mid \Theta_Y = \theta_2]$, for all $i \in [n]$ and for any $\theta_1 \leq \theta_2$, $\theta_1, \theta_2 \in \mathscr{T}$,

(4) $[X_i(\Theta_X) \mid \Theta_X = \theta] \leq_{disp} [Y_i(\Theta_Y) \mid \Theta_Y = \theta]$ for all $i \in [n]$ and for any $\theta \in \mathscr{T}$,

(5) $\Theta_X \leq_{st} \Theta_Y$.

Also, assume that $[X_i(\Theta_X) \mid \Theta_X = \theta]$ *and* $[Y_i(\Theta_Y) \mid \Theta_Y = \theta]$ *satisfy the properties (P1) and (P2) for any* $i \in [n]$ *and* $\theta \in \mathscr{T}$. *Then, for any weighting vector* $\vec{w} \in [0,1]^n$,

$$\sum_{i=1}^{n} w_i \left(X_i(\Theta_X) - \sum_{j=1}^{n} w_j X_j(\Theta_X) \right)^2 \leq_{st} \sum_{i=1}^{n} w_i \left(Y_i(\Theta_Y) - \sum_{j=1}^{n} w_j Y_j(\Theta_Y) \right)^2,$$

and

$$\sum_{i=1}^{n} w_i \left(X_{(i)}(\Theta_X) - \sum_{j=1}^{n} w_j X_{(j)}(\Theta_X) \right)^2 \leq_{st} \sum_{i=1}^{n} w_i \left(Y_{(i)}(\Theta_Y) - \sum_{j=1}^{n} w_j Y_{(j)}(\Theta_Y) \right)^2.$$

Proof: Consider the case $[X_i(\Theta_X) \mid \Theta_X = \theta_1] \leq_{disp} [X_i(\Theta_X) \mid \Theta_X = \theta_2]$ for any $\theta_1 \leq \theta_2$ in assumption (3) (the case $[Y_i(\Theta_Y) \mid \Theta_Y = \theta_1] \leq_{disp} [Y_i(\Theta_Y) \mid \Theta_Y = \theta_2]$ in (3) has a similar proof).

By assumptions (1), (2) and (3), applying Theorem 5.29 one has that

$$\left[\sum_{i=1}^{n} w_i \left(X_i(\Theta_X) - \sum_{j=1}^{n} w_j X_j(\Theta_X) \right)^2 \middle| \Theta_X = \theta_1 \right] \leq_{st}$$
$$\leq_{st} \left[\sum_{i=1}^{n} w_i \left(X_i(\Theta_X) - \sum_{j=1}^{n} w_j X_j(\Theta_X) \right)^2 \middle| \Theta_X = \theta_2 \right],$$

for any $\theta_1, \theta_2 \in \mathscr{T}$ such that $\theta_1 \leq \theta_2$. Using again assumption (5) and Proposition 2.105,

$$\sum_{i=1}^{n} w_i \left(X_i(\Theta_X) - \sum_{j=1}^{n} w_j X_j(\Theta_X) \right)^2 \leq_{st} \sum_{i=1}^{n} w_i \left(X_i(\Theta_Y) - \sum_{j=1}^{n} w_j X_j(\Theta_Y) \right)^2. \quad (5.4)$$

By assumptions (1), (2) and (4), applying Theorem 5.29 one has that

$$\left[\sum_{i=1}^{n} w_i \left(X_i(\Theta_X) - \sum_{j=1}^{n} w_j X_j(\Theta_X) \right)^2 \middle| \Theta_X = \theta \right] \leq_{st}$$
$$\leq_{st} \left[\sum_{i=1}^{n} w_i \left(Y_i(\Theta_Y) - \sum_{j=1}^{n} w_j Y_j(\Theta_Y) \right)^2 \middle| \Theta_Y = \theta \right],$$

for any $\theta \in \mathscr{T}$. Then, by which in turn, by assumption (5) and Theorem 2.105, implies

$$\sum_{i=1}^{n} w_i \left(X_i(\Theta_Y) - \sum_{j=1}^{n} w_j X_j(\Theta_Y) \right)^2 \leq_{st} \sum_{i=1}^{n} w_i \left(Y_i(\Theta_Y) - \sum_{j=1}^{n} w_j Y_j(\Theta_Y) \right)^2. \quad (5.5)$$

Therefore, the first statement follows from (5.4), (5.5) and the transitivity of the usual stochastic order. For the second inequality the proof is the same but using Theorem 5.30 instead of Theorem 5.29. ∎

The proof of the following statement is similar to the previous one and, therefore, it is omitted.

Theorem 5.32 *[38] Let $\vec{X}(\Theta_X)$ and $\vec{Y}(\Theta_Y)$ be two random vectors depending on two random parameters Θ_X and Θ_Y with the same support $\mathscr{T}$ such that:*

(1) Each one of $[\vec{X}(\Theta_X) \mid \Theta_X = \theta]$ and $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta]$ is a random vector with identically distributed marginals for any $\theta \in \mathscr{T}$,

(2) $[\vec{X}(\Theta_X) \mid \Theta_X = \theta_1]$, $[\vec{X}(\Theta_X) \mid \Theta_X = \theta_2]$, $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta_1]$ and $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta_2]$ have the same copula for any $\theta_1, \theta_2 \in \mathscr{T}$,

(3) $[X_i(\Theta_X) \mid \Theta_X = \theta_1] \geq_{disp} [X_i(\Theta_X) \mid \Theta_X = \theta_2]$, or $[Y_i(\Theta_Y) \mid \Theta_Y = \theta_1] \geq_{disp} [Y_i(\Theta_Y) \mid \Theta_Y = \theta_2]$, for all $i \in [n]$ and for any $\theta_1 \leq \theta_2$, $\theta_1, \theta_2 \in \mathscr{T}$,

(4) $[X_i(\Theta_X) \mid \Theta_X = \theta] \geq_{disp} [Y_i(\Theta_Y) \mid \Theta_Y = \theta]$ for all $i \in [n]$ and for any $\theta \in \mathscr{T}$,

(5) $\Theta_X \leq_{st} \Theta_Y$.

Then, for any weighting vector $\vec{w} \in [0,1]^n$ it holds that

$$\sum_{i=1}^{n} w_i \left(X_i(\Theta_X) - \sum_{j=1}^{n} w_j X_j(\Theta_X) \right)^2 \geq_{st} \sum_{i=1}^{n} w_i \left(Y_i(\Theta_Y) - \sum_{j=1}^{n} w_j Y_j(\Theta_Y) \right)^2, \quad (5.6)$$

and

$$\sum_{i=1}^{n} w_i \left(X(i)(\Theta_X) - \sum_{j=1}^{n} w_j X_{(j)}(\Theta_X) \right)^2 \geq_{st} \sum_{i=1}^{n} w_i \left(Y_{(i)}(\Theta_Y) - \sum_{j=1}^{n} w_j Y_{(j)}(\Theta_Y) \right)^2. \quad (5.7)$$

The following is an example where the stochastic order between the weighted variability measures is guaranteed by the application of Theorem 5.32.

Example 5.33 *[38] Let $\vec{X}(\Theta_X)$ and $\vec{Y}(\Theta_Y)$ be two vectors described by two multivariate frailty models [244] that have conditionally independent exponentially distributed underlying lifetimes with rate $\lambda = 1$ and frailties Θ_X and Θ_Y, respectively, with $\Theta_X \sim \Gamma(1,\beta_X)$ and $\Theta_Y \sim \Gamma(1,\beta_Y)$.*

That is, let $\vec{X}(\Theta_X)$ and $\vec{Y}(\Theta_Y)$ have joint survival functions

$$\overline{F}_{\vec{X}(\Theta_X)}(t_1,\ldots,t_n) = E_{\Theta_X}\left[\prod_{i=1}^{n}\left(e^{-t_i}\right)^{\Theta_X}\right] = \overline{W}_{\vec{X}}\left(\sum_{i=1}^{n} t_i\right),$$
$$\overline{F}_{\vec{Y}(\Theta_Y)}(t_1,\ldots,t_n) = E_{\Theta_Y}\left[\prod_{i=1}^{n}\left(e^{-t_i}\right)^{\Theta_Y}\right] = \overline{W}_{\vec{Y}}\left(\sum_{i=1}^{n} t_i\right),$$

where $\overline{W}_{\vec{X}}(t) = E_{\Theta_X}[e^{-t\Theta_X}] = (1+t)^{-\beta_X}$ and $\overline{W}_{\vec{Y}}(t) = E_{\Theta_Y}[e^{-t\Theta_Y}] = (1+t)^{-\beta_Y}$.

Assume $\beta_X < \beta_Y$. Conditions (1-5) in Theorem 5.32 are fulfilled because of the following facts:

(1) Both $[\vec{X}(\Theta_X) \mid \Theta_X = \theta]$ and $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta]$ are random vectors of independent and exponentially distributed variables with the same rate θ,

(2) $[\vec{X}(\Theta_X) \mid \Theta_X = \theta]$ and $[\vec{Y}(\Theta_Y) \mid \Theta_Y = \theta]$ have the independence copula for any θ,

(3) $[X_i(\Theta_X) \mid \Theta_X = \theta_1] \geq_{disp} [X_i(\Theta_X) \mid \Theta_X = \theta_2]$ for any $\theta_1 \leq \theta_2$, since both are vectors of independent and exponentially distributed variables with rates θ_1 and θ_2, respectively, and it is easy to check that $Z_1 \geq_{disp} Z_2$ whenever $Z_i \sim Exp(\theta_i)$ and $\theta_1 \leq \theta_2$,

(4) The variables $[X_i(\Theta_X) \mid \Theta_X = \theta]$ and $[Y_i(\Theta_Y) \mid \Theta_Y = \theta]$ have the same distribution for any θ,

(5) $\Theta_X \leq_{st} \Theta_Y$ by properties of the gamma distribution.

The properties (P1) and (P2) for $[X_i(\Theta_X) \mid \Theta_X = \theta]$ and $[Y_i(\Theta_Y) \mid \Theta_Y = \theta]$ are clearly satisfied for any $\theta \in \mathscr{T}$. Thus, Theorem 5.32 can be applied, and, for

any weighting vector $\vec{w} \in [0,1]$, the stochastic inequalities in Equations 5.6 and 5.7 hold.

Note that the vectors $\vec{X}(\Theta_X)$ and $\vec{Y}(\Theta_Y)$ have different marginal survival functions, being $\overline{F}_{X_i(\Theta_X)}(t) = (1+t)^{-\beta_X}$ and $\overline{F}_{Y_i(\Theta_Y)}(t) = (1+t)^{-\beta_Y}$ with $t \geq 0$, and different copulas. In fact, $\vec{X}(\Theta_X)$ and $\vec{Y}(\Theta_Y)$ are special cases of the time transformed exponential model (TTE), in which the considered survival copulas are $\widehat{C}_{\vec{X}}(u_1,\ldots,u_n) = \overline{W}_{\vec{X}}\left(\sum_{i=1}^{n} \overline{W}_{\vec{X}}^{-1}(u_i)\right)$ and $\widehat{C}_{\vec{Y}}(u_1,\ldots,u_n) = \overline{W}_{\vec{Y}}\left(\sum_{i=1}^{n} \overline{W}_{\vec{Y}}^{-1}(u_i)\right)$. For more details in this regard, the reader is referred to [244].

5.2.2 Conditions for the increasing convex order

As introduced in Section 2.3.5.2, other important variability orders are the convex and componentwise convex orders. Similar results to those for the dispersive order can be proved using these orders, but with the weaker increasing convex order between variability estimators.

The first statement is similar to Theorem 5.27. Denote as $\vec{X}_D = (X_i - X_j, i, j \in [n])$ the random vector consisting of the difference between the possible combinations of the components of $\vec{X}$.

Theorem 5.34 *[38] Let $\vec{X}$ and $\vec{Y}$ be two random vectors such that $\vec{X} \leq_{ccx} \vec{Y}$. Then,*

$$\vec{X}_D \leq_{cx} \vec{Y}_D.$$

Proof: By Theorem 3.2 in [245], it suffices to prove that $E[\varphi(\vec{X}_D)] \leq E[\varphi(\vec{Y}_D)]$ for any twice differentiable convex function $\varphi : \mathbb{R}^{n^2} \to \mathbb{R}$. Thus, let φ be any twice differentiable convex function. Denote as φ_i and φ_{ij}, with $i, j \in [n]$, its first and second derivatives. In addition, denote as H its Hessian, which is positive semi-definite. Now, consider the function $\phi : \mathbb{R}^n \to \mathbb{R}^{n^2}$ defined as $\phi(\vec{x}) = \{(x_i - x_j)\}_{i,j\in[n]}$, and compute $\frac{\partial^2 \varphi \circ \phi}{\partial x_k^2}$ with $k \in [n]$.

Firstly, denote as I^+ the set of indices such that $i \in I^+$ if and only if $(\phi(\vec{x}))_i = x_k - x_j$ with $j \in [n]$. Similarly, denote as I^- the set of indices such that $i \in I^-$ if and

only if $(\phi(\vec{x}))_i = x_j - x_k$ with $j \in [n]$. Then:

$$\frac{\partial \varphi \circ \phi}{\partial x_k} = \sum_{i \in I^+} \varphi_i \circ \phi - \sum_{i \in I^-} \varphi_i \circ \phi,$$

$$\frac{\partial^2 \varphi \circ \phi}{\partial x_k^2} = \sum_{i,j \in I^+} \varphi_{ij} \circ \phi + \sum_{i,j \in I^-} \varphi_{ij} \circ \phi - \sum_{i \in I^+, j \in I^-} \varphi_{ij} \circ \phi - \sum_{i \in I^-, j \in I^+} \varphi_{ij} \circ \phi.$$

Now, consider the vector $\vec{s}$ of dimension n^2 defined such that $s_i = 1$ if $i \in I^+$, $s_i = -1$ if $i \in I^-$ and $s_i = 0$ if $i \notin I^+ \cup I^-$. The last expression of the second derivative is equivalent to $\frac{\partial^2 \varphi \circ \phi}{\partial x_k^2} = \vec{s}^t (H \circ \phi) \vec{s}$. Since H is positive semi-definite, $\frac{\partial^2 \varphi \circ \phi}{\partial x_k^2} \geq 0$. This holds for any $k \in [n]$, thus $\varphi \circ \phi$ is a componentwise convex function.

Then, since $\vec{X} \leq_{ccx} \vec{Y}$, it holds $E[\varphi \circ \phi(\vec{X})] \leq E[\varphi \circ \phi(\vec{Y})]$ and $E[\varphi(\vec{X}_D)] \leq E[\varphi(\vec{Y}_D)]$ for any twice differentiable convex function φ. ∎

Unfortunately, from last result does not follow a statement similar to Corollary 5.28 but having the convex order as an assumption. The reason is that the composition between two convex functions (or between an increasing convex and a convex function) is not necessarily convex (or increasing convex). However, conditions to order sample variances in the increasing convex order, under the weaker assumption of convex order rather than the componentwise convex order, are provided in the following statement. On the contrary as in Theorem 5.29, no additional conditions on the same marginal distribution or the same copula for the random vectors are required. Moreover, the result involves two possible different weighting vectors for the pondering of the mean and the squared differences.

Theorem 5.35 *[38] Let $\vec{X}$ and $\vec{Y}$ be two random vectors such that $\vec{X} \leq_{cx} \vec{Y}$. Then, for any pair of weighting vectors $\vec{w} \in [0,1]^n$ and $\vec{v} \in [0,1]^n$ it holds*

$$\sum_{i=1}^{n} w_i \left(X_i - \sum_{j=1}^{n} v_j X_j \right)^2 \leq_{icx} \sum_{i=1}^{n} w_i \left(Y_i - \sum_{j=1}^{n} v_j Y_j \right)^2.$$

Proof: Since $\vec{X} \leq_{cx} \vec{Y}$, there exist two random vectors $\widehat{\vec{X}}$ and $\widehat{\vec{Y}}$ defined in the same probability space such that $\vec{X} =_{st} \widehat{\vec{X}}$, $\vec{Y} =_{st} \widehat{\vec{Y}}$ and $E[\widehat{\vec{Y}} | \widehat{\vec{X}}] =_{a.s.} \widehat{\vec{X}}$ (see Theorem 2.117).

Define the function $h : \mathbb{R}^+ \to \mathbb{R}^+$ as

$$h(\lambda) = E\left[\sum_{i=1}^{n} w_i \left(\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j\right)^2 \middle| \sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j\right)^2 = \lambda\right].$$

Consider a vector $\vec{x} \in \mathbb{R}^n$ such that $\sum_{i=1}^{n} w_i \left(x_i - \sum_{j=1}^{n} v_j x_j\right)^2 = \lambda$. If $\widehat{\vec{X}} = \vec{x}$, then $\widehat{\vec{Y}}$ can be expressed as $\widehat{\vec{Y}} = \vec{x} + \vec{\varepsilon}$, with $\vec{\varepsilon}$ being a random vector with a null mean vector. Therefore,

$$E\left[\sum_{i=1}^{n} w_i \left(\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j\right)^2 \middle| \widehat{\vec{X}} = \vec{x}\right] = E\left[\sum_{i=1}^{n} w_i \left(x_i - \sum_{j=1}^{n} v_j x_j + \varepsilon_i - \sum_{j=1}^{n} v_j \varepsilon_j\right)^2\right].$$

Expanding the square in the right side of the equation, each summand is of the form

$$E\left[\left(x_i - \sum_{j=1}^{n} v_j x_j\right)^2 + \left(\varepsilon_i - \sum_{j=1}^{n} v_j \varepsilon_j\right)^2 + 2\left(x_i - \sum_{j=1}^{n} v_j x_j\right)\left(\varepsilon_i - \sum_{j=1}^{n} v_j \varepsilon_j\right)\right],$$

for any $i \in [n]$. Then, using that the expectation of $\vec{\varepsilon}$ is $\vec{0}$, it holds that

$$E\left[\sum_{i=1}^{n} w_i \left(\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j\right)^2 \middle| \widehat{\vec{X}} = \vec{x}\right] =$$

$$= E\left[\sum_{i=1}^{n} w_i \left(\left(x_i - \sum_{j=1}^{n} v_j x_j\right)^2 + \left(\varepsilon_i - \sum_{j=1}^{n} v_j \varepsilon_j\right)^2\right)\right] \geq$$

$$\geq \sum_{i=1}^{n} w_i \left(x_i - \sum_{j=1}^{n} v_j x_j\right)^2 = \lambda.$$

Define the set $C_\lambda = \left\{\vec{x} \in \mathbb{R}^n : \sum_{i=1}^{n} w_i \left(x_i - \sum_{j=1}^{n} v_j x_j\right)^2 = \lambda\right\}$ and denote as f the density function of the conditional distribution of $\widehat{\vec{X}}$ given $\widehat{\vec{X}} \in C_\lambda$. Using it,

$$h(\lambda) = \int_{\vec{x} \in C_\lambda} E\left[\sum_{i=1}^{n} w_i \left(\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j\right)^2 \middle| \widehat{\vec{X}} = \vec{x}\right] f(\vec{x}) d\vec{x} \geq \int_{\vec{x} \in C_\lambda} \lambda f(\vec{x}) d\vec{x} = \lambda.$$

Using the latter inequality,

$$h\left(\sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j\right)^2\right) \geq_{a.s} \sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j\right)^2.$$

Then, it is concluded that

$$E\left[\sum_{i=1}^{n} w_i \left(\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j\right)^2 \middle| \sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j\right)^2\right] \geq \sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j\right)^2.$$

In addition, since

$$\sum_{i=1}^{n} w_i \left(\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j\right)^2 =_{st} \sum_{i=1}^{n} w_i \left(X_i - \sum_{j=1}^{n} v_j X_j\right)^2,$$

and

$$\sum_{i=1}^{n} w_i \left(\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j\right)^2 =_{st} \sum_{i=1}^{n} w_i \left(Y_i - \sum_{j=1}^{n} v_j Y_j\right)^2,$$

applying Theorem 2.109 it holds that

$$\sum_{i=1}^{n} w_i \left(X_i - \sum_{j=1}^{n} v_j X_j\right)^2 \leq_{icx} \sum_{i=1}^{n} w_i \left(Y_i - \sum_{j=1}^{n} v_j Y_j\right)^2.$$

■

Unfortunately, a similar result cannot be stated when working with the order statistics, since the convex order is not preserved when the ordering is applied. However, it is possible to prove a similar result dealing with the Gini mean difference.

Theorem 5.36 *[38] Let $\vec{X}$ and $\vec{Y}$ be two random vectors such that $\vec{X} \leq_{cx} \vec{Y}$. Then,*

$$\frac{1}{n(n-1)} \sum_{i,j=1}^{n} \left|X_i - X_j\right| \leq_{icx} \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \left|Y_i - Y_j\right|.$$

Proof: Since $\vec{X} \leq_{cx} \vec{Y}$, then there exist two random vectors $\widehat{\vec{X}}$ and $\widehat{\vec{Y}}$ defined in the same probability space such that $\vec{X} =_{st} \widehat{\vec{X}}$, $\vec{Y} =_{st} \widehat{\vec{Y}}$ and $E[\widehat{\vec{Y}}|\widehat{\vec{X}}] =_{a.s.} \widehat{\vec{X}}$ (see Theorem 2.117).

Define the function $h : \mathbb{R}^+ \to \mathbb{R}^+$ as

$$h(\lambda) = E\left[\frac{1}{n(n-1)}\sum_{i,j=1}^{n}|\widehat{Y}_i - \sum_{j=1}^{n} v_j \widehat{Y}_j| \,\middle|\, \frac{1}{n(n-1)}\sum_{i,j=1}^{n}|\widehat{X}_i - \sum_{j=1}^{n} v_j \widehat{X}_j| = \lambda\right].$$

Consider a vector $\vec{x} \in \mathbb{R}^n$ such that $\frac{1}{n(n-1)}\sum_{i,j=1}^{n}\left|x_i - x_j\right| = \lambda$. If $\widehat{\vec{X}} = \vec{x}$, $\widehat{\vec{Y}}$ can be expressed as $\widehat{\vec{Y}} = \vec{x} + \vec{\varepsilon}$, with $\vec{\varepsilon}$ being a random vector with a null mean vector. Then,

$$E\left[\frac{1}{n(n-1)}\sum_{i,j=1}^{n}\left|Y_i - Y_j\right| \,\middle|\, \widehat{\vec{X}} = \vec{x}\right] = E\left[\frac{1}{n(n-1)}\sum_{i,j=1}^{n}\left|x_i + \varepsilon_i - x_j - \varepsilon_j\right|\right] =$$
$$= \frac{1}{n(n-1)}\sum_{i,j=1}^{n} E\left[\left|x_i + \varepsilon_i - x_j - \varepsilon_j\right|\right]$$

Focusing on each of the terms, using that the expectation of the absolute value is always greater than or equal to the expectation of the original random variable, the following inequality can be computed

$$\left(E\left[\left|x_i + \varepsilon_i - x_j - \varepsilon_j\right|\right]\right)^2 \geq \left(E\left[x_i + \varepsilon_i - x_j - \varepsilon_j\right]\right)^2 = \left(x_i - x_j\right)^2 = \left|x_i - x_j\right|^2.$$

Then, one has that $E\left[\left|x_i + \varepsilon_i - x_j - \varepsilon_j\right|\right] \geq \left|x_i - x_j\right|$ and, therefore,

$$\frac{1}{n(n-1)}\sum_{i,j=1}^{n} E\left[\left|x_i + \varepsilon_i - x_j - \varepsilon_j\right|\right] \geq \frac{1}{n(n-1)}\sum_{i,j=1}^{n}\left|x_i - x_j\right| = \lambda.$$

For the rest of the proof, proceed analogously as in Theorem 5.35. ∎

As a consequence of the latter two results, the following particular cases hold trivially.

Corollary 5.37 *[38] Let $\vec{X}$ and $\vec{Y}$ be two vectors with, respectively, independent and identically distributed components. If $X_1 \leq_{cx} Y_1$, then $\hat{\sigma}^2(\vec{X}) \leq_{icx} \hat{\sigma}^2(\vec{Y})$ and $G(\vec{X}) \leq_{icx} G(\vec{Y})$.*

Notice that the previous results are also fulfilled for the componentwise convex order, since it implies the convex order. In Figure 5.2, a summary of the main results of the latter two sections is provided.

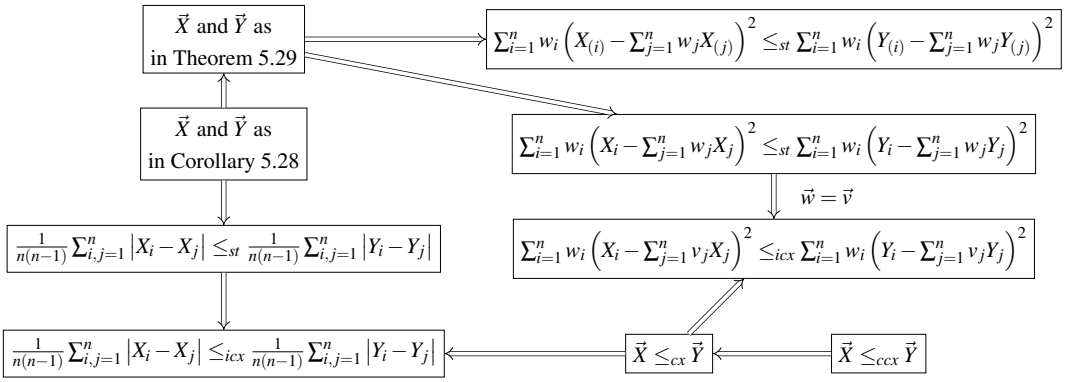

Figure 5.2: Relationship between the stochastic comparison of vectors and the stochastic ordering of the variability measures [38].

5.3 Variability measures and dependence orders

In the latter section, the results focused on the location comparison of variability measures when the random vectors were ordered by means of a variability stochastic order. However, the dependence between the components of the random sample also plays a role in the value of the variability measures.

Consider a random vector $\vec{X}$ such that the variance of each component is the same and the covariance between each pair of components is also constant. Then, the expectation of the sample variance is

$$E[\hat{\sigma}^2(\vec{X})] = \mathrm{Var}(X_1) - \mathrm{Cov}(X_1, X_2).$$

Therefore, $E[\hat{\sigma}^2(\vec{X})]$ depends on the correlation between the variables. Moreover, it is also clear that the expectation increases as the correlation between the components of the random sample decreases.

Recall that, as explained in the latter section, computing the minimum of a penalty function leads to a variability measure (the value of the minimum) and an aggregation function (the point in which the minimum is attained). Then, intuitively, when working with random vectors with strong positive dependence, one can expect smaller values for penalty functions when evaluated in its minimum. Regarding Statistics, positive dependence introduces a negative bias, for instance, in the estimation of the variance by means of the sample variance.

However, although the intuition is clear, there is a lack of proper mathematical results in this regard in the literature. In the following sections, results of the form $\vec{X} \leq_{dst} \vec{Y} \implies V(\vec{X}) \geq_{lst} V(\vec{Y})$ are studied, where $\leq_{dst}$ is a dependence stochastic order, $\leq_{lst}$ a location stochastic order and V is a variability measure.

5.3.1 Conditions for the increasing convex order

As explained above, positive dependence seems to lead to smaller values of the variability measures. The most natural result that one could expect in this regard is that, if $\vec{X} \leq_{sm} \vec{Y}$ (see Definition 2.118), then $V(\vec{X}) \geq_{st} V(\vec{Y})$. In addition, another reasonable result could be that $\vec{X} \leq_{PQD} \vec{Y}$ (see Definition 2.119) implies $V(\vec{X}) \geq_{st} V(\vec{Y})$. The following counterexample shows that, in general, the latter implications are not verified.

Example 5.38 *[40] Let $\vec{Y} = (Y_1, Y_2)$ be a random vector assuming values in $[3]^2$ and such that $P[(Y_1, Y_2) = (i, j)] = 1/9$ for all $(i, j) \in [3]^2$. Let $\vec{X} = (X_1, X_2)$ be another random vector assuming values in $[3]^2$ and such that*

$$\begin{aligned}
P[(X_1, X_2) = (i, j)] &= 14/90 \text{ if } (i, j) \in \{(1,3),(3,1)\}, \\
P[(X_1, X_2) = (i, j)] &= 6/90 \text{ if } (i, j) \in \{(1,2),(2,1),(3,2),(2,3)\}, \\
P[(X_1, X_2) = (i, j)] &= 10/90 \text{ if } (i, j) \in \{(3,3),(1,1)\}, \\
P[(X_1, X_2) = (i, j)] &= 18/90 \text{ if } (i, j) = (2,2).
\end{aligned}$$

It is easy to verify that the conditions for the PQD are satisfied, so $\vec{X} \leq_{PQD} \vec{Y}$. Simple computations show that the absolute differences $|X_1 - X_2|$ and $|Y_1 - Y_2|$ take as values 0, 1 and 2 with probabilities, respectively, $\frac{38}{90}$, $\frac{24}{90}$ and $\frac{28}{90}$ (for $|X_1 - X_2|$) and $\frac{30}{90}$, $\frac{40}{90}$ and $\frac{20}{90}$ (for $|Y_1 - Y_2|$). Noticing that $P(|X_1 - X_2| \leq 0) = \frac{38}{90} > P(|Y_1 - Y_2| \leq 0) = \frac{30}{90}$ and $P(|X_1 - X_2| \leq 1) = \frac{62}{90} < P(|Y_1 - Y_2| \leq 1) = \frac{70}{90}$. Thus, $|X_1 - X_2|$ and $|Y_1 - Y_2|$ are not ordered in usual stochastic order, and as a consequence neither $\vec{X}_{AD}$ and $\vec{Y}_{AD}$ are ordered in the usual stochastic order. It is easy to see that the variability estimators mentioned in previous sections are also not comparable according to the usual stochastic order, since they are increasing functions of $\vec{X}_{AD}$ and $\vec{Y}_{AD}$.

Notice that for bivariate random vectors, $\vec{X} \leq_{sm} \vec{Y} \iff \vec{X} \leq_{PQD} \vec{Y}$, so the same holds for the supermodular order.

In this example, it is also interesting to observe that, apart from the uniform distributions (over $[3]$) of the marginals, $\vec{X}$ and $\vec{Y}$ are exchangeable. Therefore, they are well representative of possible random samples from two random quantities X and Y.

However, other comparisons follow from the supermodular order between $\vec{X}$ and $\vec{Y}$. Recall that $\vec{X}_D$ denotes the random vector consisting of the differences of the components of $\vec{X}$.

Theorem 5.39 *[40] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then,*

$$\vec{X} \leq_{sm} \vec{Y} \implies \vec{X}_D \geq_{sym-cx} \vec{Y}_D.$$

Proof: By Theorem 3.2 in [245], it suffices to prove that $E[\varphi(\vec{X}_D)] \geq E[\varphi(\vec{Y}_D)]$ for any twice differentiable symmetric convex function $\varphi : \mathbb{R}^{n^2} \to \mathbb{R}$. Let φ be any twice differentiable symmetric convex function. Denote as φ_i and φ_{ij}, with $i, j \in [n^2]$, its first and second derivatives. In addition, denote as H its Hessian, which, since φ is convex, is positive semi-definite. Now, consider the function $\phi : \mathbb{R}^n \to \mathbb{R}^{n^2}$ defined as $\phi(\vec{x}) = \vec{x}_D$, and compute $\frac{\partial^2 \varphi \circ \phi}{\partial x_i x_j}$ for $i, j \in [n]$ and $i \neq j$.

Fix any $i \in [n]$ and denote as $I_I^+ \subseteq [n^2]$ the set of indices such that $k \in I^+$ if and only if $(\phi(\vec{x}))_k = x_i - x_s$ with $s \in [n]$ such that $s \neq i$. Similarly, denote as I^- the set of indices such that $i \in I^-$ if and only if $(\phi(\vec{x}))_k = x_s - x_i$ with $s \in [n]$ such that $s \neq i$. Then,

$$\frac{\partial \varphi \circ \phi}{\partial x_i}(\vec{x}) = \sum_{k \in I_i^+} \varphi_i \circ \phi - \sum_{k \in I_i^-} \varphi_i \circ \phi,$$

and

$$\frac{\partial^2 \varphi \circ \phi}{\partial x_i \partial x_j} = \sum_{k \in I_i^+ l \in I_j^+} \varphi_{ij} \circ \phi + \sum_{k \in I_i^- l \in I_j^-} \varphi_{ij} \circ \phi - \sum_{k \in I_i^+ l \in I_j^-} \varphi_{ij} \circ \phi - \sum_{k \in I_i^- l \in I_j^+} \varphi_{ij} \circ \phi.$$

Since φ is symmetric, then $\varphi_{ij} = \varphi_{i'j'}$ for any $i, j, i'j' \in [n^2]$ such that $i \neq j$ and $i' \neq j'$ and $\varphi_{ii} = \varphi_{jj}$ for any $i, j \in [n^2]$. Notice that $I_i^+ \cap I_j^+ = I_i^- \cap I_j^- - = \emptyset$ if $i \neq j$. In addition, $I_i^+ \cap I_j^-$ and $I_i^- \cap I_j^+$ only have one element each (associated with the

component such that ϕ equals $x_i - x_j$ and $x_j - x_i$). Then,

$$\begin{aligned}\frac{\partial^2 \varphi \circ \phi}{\partial x_i \partial x_j}(\vec{x}) &= 2n^2 \varphi_{12} \circ \phi(\vec{x}) - 2(n^2-2)\varphi_{12} \circ \phi(\vec{x}) - 2\varphi_{11} \circ \phi(\vec{x}) \\ &= -2\varphi_{11} \circ \phi(\vec{x}) + 2\varphi_{12} \circ \phi(\vec{x}) \\ &= -\varphi_{11} \circ \phi(\vec{x}) - \varphi_{22} \circ \phi(\vec{x}) + \varphi_{12} \circ \phi(\vec{x}) + \varphi_{21} \circ \phi(\vec{x}).\end{aligned}$$

The latter expression is negative since

$$\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \varphi_{11} \circ \phi(\vec{x}) & \varphi_{12} \circ \phi(\vec{x}) \\ \varphi_{21} \circ \phi(\vec{x}) & \varphi_{22} \circ \phi(\vec{x}) \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \geq 0,$$

where the involved matrix is a diagonal submatrix of $H \circ \phi$, which is positive semi-definite.

Then, $\varphi \circ \phi$ is submodular while $-\varphi \circ \phi$ is supermodular. Applying that $\vec{X} \leq_{sm} \vec{Y}$, one has $E[-\varphi \circ \phi(\vec{X})] \leq E[-\varphi \circ \phi(\vec{Y})]$ and, therefore, $E[\varphi(\vec{X}_D)] \geq E[\varphi(\vec{Y}_D)]$ for any twice differentiable symmetric convex function φ. ■

The latter result is a multivariate generalization of Theorem 4 in [246]. As an immediate consequence, one gets the following statement regarding vectors containing the absolute differences.

Corollary 5.40 *[40] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then,*

$$\vec{X} \leq_{sm} \vec{Y} \implies \vec{X}_{AD} \geq_{sym-icx} \vec{Y}_{AD}.$$

Proof: Let $\phi : \mathbb{R}^{n^2} \to \mathbb{R}$ be any symmetric increasing convex function and let $g : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ be defined as $g(x_1, \ldots, x_{n^2}) = (g_1(\vec{x}), \ldots, g_{n^2}(\vec{x})) = (|x_1|, \ldots, |x_{n^2}|)$. Note that the functions $g_1, \ldots, g_{n^2}$ are convex.

Define $h : \mathbb{R}^{n^2} \to \mathbb{R}$ as $h = \phi \circ g$. By the properties of the composition of functions listed in page 86 of [69] it holds that h is convex. Moreover, it is symmetric. Thus, applying Theorem 5.39 one has,

$$\begin{aligned}&E\left[\phi\left(\vec{X}_{AD}\right)\right] = E\left[\phi\left(g\left(\vec{X}_D\right)\right)\right] = E\left[h\left(\vec{X}_D\right)\right] \geq \\ &\geq E\left[h\left(\vec{Y}_D\right)\right] = E\left[\phi\left(g\left(\vec{Y}_D\right)\right)\right] = E\left[\phi\left(\vec{Y}_{AD}\right)\right].\end{aligned}$$

and the result holds. ∎

It must be pointed out that Corollary 5.40 is a multivariate generalization of Theorem 13 in [259]. Also, from the corollary easily follow conditions to let some variability estimators be ordered in the increasing convex order.

Corollary 5.41 *[40] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. If $\vec{X} \leq_{sm} \vec{Y}$, then it holds $\hat{\sigma}^2(\vec{X}) \geq_{icx} \hat{\sigma}^2(\vec{Y})$, $G(\vec{X}) \geq_{icx} G(\vec{Y})$ and $R(\vec{X}) \geq_{icx} R(\vec{Y})$.*

Proof: Observe that $\hat{\sigma}^2(\vec{X}) = h_1(\vec{X}_{AD})$, $G(\vec{Y}) = h_2(\vec{X}_{AD})$ and $R(\vec{X}) = h_3(\vec{X}_{AD})$ with

$$h_1(\vec{x}) = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2,$$

$$h_2(\vec{x}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i,$$

and

$$h_3(\vec{x}) = \max_{i \in [n]} \{x_i\}.$$

Notice that h_1, h_2 and h_3 are symmetric, increasing and convex when $x_i \geq 0$ for any $i \in [n]$. Thus, applying Corollary 5.40 and recalling that the composition of increasing convex functions is increasing convex,

$$\vec{X} \leq_{sm} \vec{Y} \implies \vec{X}_{AD} \geq_{sym-icx} \vec{Y}_{AD} \implies h_i(\vec{X}_{AD}) \geq_{icx} h_i(\vec{X}_{AD}),$$

for any $i \in \{1,2,3\}$. ∎

Regarding the stochastic comparison between the sample ranges $R(\vec{X})$ and $R(\vec{Y})$, Corollary 5.41 is an improvement of Proposition 3.2 in [82], where it is shown that if $(X_1, X_2) \leq_{sm} (Y_1, Y_2)$ then $E[|X_1 - X_2|] \geq E[|Y_1 - Y_2|]$ The improvement is in two directions: firstly it shows the validity of a stronger comparison between the two absolute differences under the same assumption, and secondly it generalizes the result to dimensions greater than 2.

In the following, two examples of the applicability of the last result in insurance and estimation are provided.

Example 5.42 *Consider an insurance company that has a number n of insureds, to which can correspond n independent and identically distributed claims $X_1,\ldots,X_n$. To these claims assign an exchangeable vector $\vec{I}^X = (I_1^X,\ldots,I_n^X)$ consisting of random variables with Bernoulli distribution describing the occurrence levels of the claims and consider the total amount of the claim $\sum_{j=1}^n I_j^X X_j$ faced by the insurance company. This model is widely considered in the literature (see [334] and references in that paper). Moreover, it is of interest to establish conditions for comparing the total amount of the claim under different assumptions. As noted in [16], the difference between the largest and the smallest claims can be considered as a criterion to evaluate the total amount of the claim. Due to the presence of the terms I_j^X and the dependence between them, an explicit analytical expression of the distribution function of $\max_{j\in[n]} I_j^X X_j - \min_{j\in[n]} I_j^X X_j$ is generally unavailable. However, bounds on this quantity in terms of the increasing convex order can be provided. Assume, for example, that $\vec{I}^Y \leq_{sm} \vec{I}^X$, where $\vec{I}^Y = (I_1^Y,\ldots,I_n^Y)$ is a vector of independent Bernoulli random variables that have the same parameters as those of $\vec{I}^X$. If the occurrences $\vec{I}^X$ have a positive dependence, which is the usual case for the occurrences of claims, by Theorem 9.A.12 in [295] one has that $(I_1^Y X_1,\ldots,I_n^Y X_n) \leq_{sm} (I_1^X X_1,\ldots,I_n^X X_n)$. Thus, from Corollary 5.41 it follows*

$$\max_{j\in[n]} I_j^Y X_j - \min_{j\in[n]} I_j^Y X_j \geq_{icx} \max_{j\in[n]} I_j^X X_j - \min_{j\in[n]} I_j^X X_j,$$

where the distribution in the left side of the inequality can be easily computed, being the components of $(I_1^Y X_1,\ldots,I_n^Y X_n)$ independent.

Example 5.43 *[40] Consider $(X_1,\ldots,X_n)$ to be a sample of observations of a uniform distribution over an unknown interval $[a,b]$. The Maximum Likelihood Estimators for the parameters a and b are, respectively, $\min(\vec{X})$ and $\max(\vec{Y})$, see [278]. In addition, the variance of a uniform distribution takes the value $\frac{1}{12}(b-a)^2$. Thus, a natural estimation of the variance is*

$$\frac{1}{12}\left(\max(\vec{X}) - \min(\vec{X})\right)^2 = \frac{1}{12} R(\vec{X})^2.$$

Consider $(Y_1,\ldots,Y_n)$ to be a new sample of observations of the same uniform distribution such that $(X_1,\ldots,X_n) \leq_{sm} (Y_1,\ldots,Y_n)$. Then, as proved in Corol-

lary 5.41, $R_n(X) \geq_{icx} R_n(Y)$ and, therefore, since $f(x) = x^2$ is increasing and convex,

$$E\left[\frac{1}{12}R_n(X)^2\right] \geq E\left[\frac{1}{12}R_n(Y)^2\right].$$

Notice that, on the contrary as in the case of the sample variance described in Section 5.2, the latter inequality of expectations cannot be achieved simply by considering the comparison of the covariances.

5.3.2 Conditions for the usual stochastic order

Example 5.38 shows that, in general, $\vec{X} \leq_{sm} \vec{Y}$ does not imply the usual stochastic order between the marginal distributions of $\vec{X}_{AD}$ and $\vec{Y}_{AD}$, i.e., does not imply $|X_i - X_j| \geq_{st} |Y_i - Y_j|$ for $i, j \in [n]$ such that $i \neq j$. However, there are cases where this stochastic inequality is actually satisfied. The following example considers bivariate exchangeable Bernoulli distributions.

Example 5.44 *[40] Let $\vec{X} = (X_1, X_2)$ and $\vec{Y} = (Y_1, Y_2)$ be two bivariate vectors having exchangeable Bernoulli distributed marginals of the same parameter $p \in (0,1)$. Consider $p^X_{00}, p^X_{11} \in [0,1]$ such that $p^X_{00} + p^X_{11} \leq 1$ and $p^X_{00}, p^X_{11} \leq p$. Let (X_1, X_2) be such that $P[(X_1, X_2) = (0,0)] = p^X_{00}$, $P[(X_1, X_2) = (1,1)] = p^X_{11}$ and the remaining probability is equally distributed over the points $(0,1)$ and $(1,0)$. Similarly, define the distribution of (Y_1, Y_2), denoting p^Y_{00} instead of p^X_{00} and p^Y_{11} instead of p^X_{11}. It is easy to verify that $\vec{X} \leq_{sm} \vec{Y}$ holds if and only if $p^X_{00} \leq p^Y_{00}$ and $p^X_{11} \leq p^Y_{11}$.*

Moreover, observe that $|X_1 - X_2|$ has Bernoulli distribution with parameter $1 - (p^X_{00} + p^X_{11})$ while $|Y_1 - Y_2|$ has Bernoulli distribution with parameter $1 - (p^Y_{00} + p^Y_{11})$. Thus, $\vec{X} \leq_{sm} \vec{Y}$ implies $|X_1 - X_2| \geq_{st} |Y_1 - Y_2|$.

Note that seeing Example 5.44 one may wonder if, in general, dealing with exchangeable multivariate Bernoulli distributions with dimensions $n > 2$, the implication $\vec{X} \leq_{sm} \vec{Y} \implies \vec{X}_{AD} \geq_{st} \vec{Y}_{AD}$ still holds true. This is not always the case, as illustrated in the following example.

Example 5.45 *[40] Observe that, for any multivariate Bernoulli random vector $\vec{X}$, the corresponding vector $\vec{X}_{AD}$ still has a multivariate Bernoulli distribution, whose*

dimension is n^2. Consider now the sum $\sum_{i,j=1}^{n}|X_i - X_j|$ of the components of $\vec{X}_{AD}$. It is easy to verify that it is equivalent to $N_X(n-N_X)$, where $N_X = \sum_{i=1}^{n} X_i$ [133].

Let now $\vec{X}$ and $\vec{Y}$ be two vectors of dimension $n = 5$ having exchangeable multivariate Bernoulli distributions whose marginally distributions are Bernoulli with parameter $p = 0.5$. Assume that $P(N_X = 2) = \frac{3}{4}$, $P(N_X = 4) = \frac{1}{4}$ and $P(N_X = k) = 0$ for any $k \in \{0,1,3,5\}$. In addition, suppose that $P(N_Y = 2) = \frac{5}{6}$, $P(N_Y = 5) = \frac{1}{6}$ and $P(N_Y = k) = 0$ for any $k \in \{0,1,3,4\}$.

Note that the vectors $\vec{X}$ and $\vec{Y}$ defined in this manner are the cases $r_8(y)$ and $r_9(y)$ of Table 3 in [129]. It is easy to verify that $N_X \leq_{cx} N_Y$, thus also $\vec{X} \leq_{sm} \vec{Y}$ (see Proposition 1 in [133]). Using the relations $\sum_{i,j=1}^{n}|X_i - X_j| = N_X(5-N_X)$ and $\sum_{i,j=1}^{n}|Y_i - Y_j| = N_Y(5-N_Y)$ one can analytically find the distributions of them, obtaining that $\sum_{i,j=1}^{n}|X_i - X_j|$ is equal 4 with probability $\frac{1}{4}$ and equal 6 with probability $\frac{3}{4}$, while $\sum_{i,j=1}^{n}|Y_i - Y_j|$ is equal 0 with probability $\frac{1}{6}$ and equal 6 with probability $\frac{5}{6}$. Their cumulative distributions cross each other, thus they are not ordered in the usual stochastic order. Therefore, and since the sum is an increasing function, $\vec{X}_{AD}$ and $\vec{Y}_{AD}$ also cannot be ordered in the usual stochastic order.

Another example of vectors $\vec{X}$ and $\vec{Y}$ such that $\vec{X} \leq_{sm} \vec{Y}$ implies $|X_i - X_j| \geq_{st} |Y_i - Y_j|$ is whenever they have a copula in the FGM family and are exchangeable.

Theorem 5.46 *[40] Let $\vec{X} = (X_1, X_2)$ and $\vec{Y} = (Y_1, Y_2)$ be two exchangeable vectors having FGM copulas. If $\vec{X} \leq_{sm} \vec{Y}$ then $|X_1 - X_2| \geq_{st} |Y_1 - Y_2|$.*

Proof: Let $\vec{X} = (X_1, X_2)$ be exchangeable and having an FGM copula with parameter θ_X. As proved in Theorem 2 in [64], it admits the representation

$$\vec{X} = (\vec{1} - \vec{I})\vec{X}_{[1]} + \vec{I}\vec{X}_{[2]}, \tag{5.8}$$

where:

- $\vec{X}_{[1]}$ is a bivariate vector whose components are independent and having the same distribution as $\min(\vec{X})$,
- $\vec{X}_{[2]}$ is a bivariate vector whose components are independent and having the same distribution as $\max(\vec{X})$,

- $\vec{I}$ is a vector having an exchangeable bivariate Bernoulli distribution such that $P(\vec{I}=(0,0))=P(\vec{I}=(1,1))=\frac{1+\theta_X}{4}$ and $P(\vec{I}=(0,1))=P(\vec{I}=(1,0))=\frac{1-\theta_X}{4}$.

Let X_1' and X_2' be independent copies of X_1 and X_2. Observe that if $\vec{I}=(0,0)$, $|X_1-X_2|$ has the same distribution as $\Delta_0=|\min(\vec{X})-\min(X_1',X_2')|$, if $\vec{I}=(1,1)$ then $|X_1-X_2|$ has the same distribution as $\Delta_2=|\max(\vec{X})-\max(X_1',X_2')|$, while if $\vec{I}=(1,0)$ or $\vec{I}=(0,1)$ then $|X_1-X_2|$ has the same distribution as $\Delta_1=|\max(\vec{X})-\min(X_1',X_2')|$.

Denote as Δ_{02} the random quantity equal to Δ_0 with probability $\frac{1}{2}$ and equal to Δ_2 with probability $\frac{1}{2}$. Trivially, $\Delta_0 \leq_{st} \Delta_1$ and $\Delta_2 \leq_{st} \Delta_1$, thus applying Theorem 2.104, $\Delta_{02} \leq_{st} \Delta_1$.

Finally, observe that, by Equation 5.8, $|X_1-X_2|$ assumes the value Δ_1 with probability $(1-\theta_X)/2$, or the value Δ_{02} with probability $(1+\theta_X)/2$.

Consider now $\vec{Y}=(Y_1,Y_2)$ a random vector having the same marginal distributions as $\vec{X}$, but with the FGM copula having parameter θ_Y. Notice that $\vec{X} \leq_{sm} \vec{Y}$ is equivalent to $\theta_X \leq \theta_Y$ since in this case, and only in this case, the inequalities in Definition 2.119 are verified, so $\vec{X} \leq_{PQD} \vec{Y}$.

Similarly, the absolute difference $|Y_1-Y_2|$ takes the value Δ_1 with probability $\frac{1-\theta_Y}{2}$ and the value Δ_{02} with probability $\frac{1+\theta_Y}{2}$. Since $\theta_X \leq \theta_Y$ and $\Delta_1 \geq_{st} \Delta_{02}$, by Proposition 2.104 it follows that $|X_1-X_2| \geq_{st} |Y_1-Y_2|$. ■

The reader is referred to Section 6 in [63] for more details on the latter representation.

Chapter 6

Aggregation of other random structures

Contents

Random variables are the most prominent and considered random structures in the literature, since they are related to data that consists on real numbers. Therefore, the last chapters have been devoted to study the aggregation of such structures. However, the concept of aggregation of random variables introduced in Definition 3.6 can be extended, among others, to the aggregation to random vectors, stochastic processes and random sets.

Each random structure has its own characteristics, such as the unboundedness or not of the intervals or a particular σ-algebra in the considered measurable space. Therefore, although there are some points in common, each case needs a different approach.

Firstly, the construction of aggregations of random elements on bounded posets

is provided in Section 6.1. The definition aims to cover many random structures such as random graphs or random elements related to ordinal scales. However, the approach does not suffice to aggregate other structures. For instance, random vectors are considered in Section 6.2, since $\mathbb{R}^n$ is not bounded. Similarly, some specific considerations have been made in Section 6.3 when dealing with (possibly infinite) sets of random variables in the aggregation of stochastic processes. Furthermore, a more involved theory for the aggregation of random sets is constructed in Section 6.4, since there is not an adequate stochastic order in this regard. Finally, some brief comments are made on the aggregation of other random structures in Section 6.5.

6.1 Aggregation of random elements on bounded posets

One of the main generalizations of aggregation functions is the one defined for bounded posets, in which the monotonicity and boundary conditions are extended using the order structure [182, 197]. In relation to data analysis, ordinal data, such as the Likert scale [180], could be of interest even if the values are not real numbers.

The here considered approach is equivalent to the one in [31] but more general, since the lattice structure is not needed. Moreover, Theorem 2.40 is used to simplify some of the proofs.

Consider $(S, \leq)$ a partially ordered set. To construct the probabilistic structure, S must be endowed with a σ-algebra. As a generator of such σ-algebra, the set of intersections between upper and lower sets will be considered. This set is, in fact, a π-system (see Definition 2.39).

Proposition 6.1 *Let $(S, \leq)$ be a partially ordered set. Then, the set $F_s \subseteq \mathscr{P}(S)$ defined as:*

$$F_s = \{A \subseteq S : A = A_L \cap A_U, A_L \text{ lower set of } (S, \leq), A_U \text{ upper set of } (S, \leq)\},$$

is a π-system.

Proof: Clearly $S = S \cap S$ and $\emptyset = \emptyset \cap \emptyset$. If $A, B \in F_s$, then $A \cap B = (A_L \cap A_U) \cap (B_L \cap B_U) = (A_L \cap B_L) \cap (A_U \cap B_U)$. Since the intersection of upper and lower sets is, respectively, an upper and a lower set, $A \cap B \in F_s$. ∎

Moreover, in [31] it is also proved that the complementary of any element in F_s can be written as a disjoint union of elements of F_s. Now, consider the σ-algebra generated by F_s, $\mathscr{F}_s = \sigma(F_s)$. The resulting pair $(S, \mathscr{F}_s)$ is a measurable space.

These poset and measurable space structures can be extended to the Cartesian product of S, $S^n = S \times \cdots \times S$. The poset $(S^n, \leq)$ is defined considering the (componentwise) lattice order on S^n using as reference the initial order $\leq$. The Cartesian product of F_s, $F_s \times \cdots \times F_s$, is also a π-system.

Proposition 6.2 *Let $(S, \leq)$ be a partially ordered set. Then, $F_s^n = F_s \times \cdots \times F_s$ is a π-system with respect to S^n.*

Proof: Since $S, \emptyset \in F_s$, $S^n = S \times \cdots \times S \in F_s^n$ and $\emptyset = \emptyset \times \cdots \times \emptyset \in F_s^n$. Let $A, B \in F_s^n$. Then:

$$(A_1 \times \cdots \times A_n) \cap (B_1 \times \cdots \times B_n) = (A_1 \cap B_1) \times \cdots \times (A_n \cap B_n).$$

For any pair $A_i \cap B_i$ with $i \in [n]$, it holds that $A_i \cap B_i \in F_s$, thus $(A_1 \cap B_1) \times \cdots \times (A_n \cap B_n) \in F_s^n$. ∎

Notice that any element of F_s^n can be written as the intersection of an upper set and a lower set of S^n noticing that

$$\begin{aligned} A_1 \times \cdots \times A_n &= (A_{L,1} \cap A_{U,1}) \times \cdots \times (A_{L,1} \cap A_{U,1}) = \\ &= (A_{L,1} \times \cdots \times A_{L,n}) \cap (A_{U,1} \times \cdots \times A_{U,n}), \end{aligned}$$

and that the Cartesian product of upper and lower sets is, respectively, an upper and a lower set.

Therefore, one can consider $\mathscr{F}_s^n = \sigma(F_s^n)$ to construct the measurable space $(S^n, \mathscr{F}_s^n)$. Consider now a bounded poset $(S, \leq, \mathbf{0}, \mathbf{1})$ and a probabilistic space (Ω, Σ, P). The following set will be considered:

$$L_S^n(\Omega) = \{\vec{X} : \Omega \to S^n \mid \vec{X} \text{ is measurable with respect to } \mathscr{F}_s^n\}.$$

For simplicity, denote $L_S^n(\Omega)$ as L_S^n and L_S^1 as L_S. To provide a way to order these random elements, a stochastic order $\leq_{so}$ should be considered. This order is expected to be transitive, reflexive and antisymmetric with respect to the equivalence relation of having the same distribution, just as in Definition 2.100. In addition, it should preserve the initial partial order for degenerate random elements. In particular, given $\vec{x},\vec{y} \in S^n$ and $\vec{X},\vec{Y} \in L_S^n$ such that $\vec{x} \leq \vec{y}$ and $P(\vec{X} = \vec{x}) = P(\vec{Y} = \vec{y}) = 1$, it is required that $\vec{X} \leq_{so} \vec{Y}$. A definition of a possible stochastic order is provided below.

Definition 6.3 *[31] Let $\vec{X},\vec{Y} \in L_S^n$ be two random elements. If $P(\vec{X} \in U) \leq P(\vec{Y} \in U)$ for every U measurable upper set of S^n, then it is said that $\vec{X}$ is smaller than or equal to $\vec{Y}$ in the usual stochastic order and it is denoted by $\vec{X} \leq_{st} \vec{Y}$.*

It is clear that the latter definition is an extension of the usual stochastic order for random vectors. However, at this point, it is unknown if it is a stochastic order or not in the sense of Definition 2.100. The next result gives an affirmative answer to this question. Recall the definition of $\mathscr{F}_s^n$ as the σ-algebra generated by the π-system F_s^n.

Proposition 6.4 *[31] The usual stochastic order introduced in Definition 6.3 is a stochastic order in the sense of Definition 2.100.*

Proof: Transitivity is straightforward to prove. Also, if $\vec{X}$ and $\vec{Y}$ have the same distribution, then the associated probabilities for upper sets are the same, so $\vec{X} =_{st} \vec{Y}$. The other implication is not immediate. Suppose that $\vec{X}$ and $\vec{Y}$ have the same distribution over upper sets.

Let $A \in F_s^n$ be an element of the π-system introduced in Proposition 6.2. Express it as the intersection of an upper and lower set (A_U and A_L) and compute its probability,

$$P(A) = P(A_L \cap A_U) = P\left(A_U \setminus \left(\overline{A_L} \cap A_U\right)\right) = P(A_U) - P\left(\overline{A_L} \cap A_U\right).$$

It can be seen that it only depends on probabilities of upper sets, since both the complementary of a lower set and the intersection of two upper sets is an upper

set. Therefore, $P(\vec{X} \in A) = P(\vec{Y} \in A)$ for any $A \in F_s^n$. Then, apply Theorem 2.40 to conclude that $\vec{X}$ and $\vec{Y}$ have the same distribution. ∎

Since it has already been proven that $\leq_{st}$ is a proper stochastic order, all the elements that are necessary to introduce the definition of aggregations of random elements on bounded posets are settled.

Definition 6.5 *Let $(S, \leq, \mathbf{0}, \mathbf{1})$ be a bounded poset. The function $A : L_S^n(\Omega) \to L_S(\Omega)$ is an aggregation of random elements on a bounded poset (with respect to $\leq_{st}$) if the following conditions are fulfilled:*

1. *For any $\vec{X}, \vec{Y} \in L_S^n$ such that $\vec{X} \leq_{st} \vec{Y}$, $A(\vec{X}) \leq_{st} A(\vec{Y})$,*
2. *If $\vec{X} =_{a.s.} \mathbf{0}\vec{1}$, then $A(\vec{X}) =_{a.s.} \mathbf{0}$,*
3. *If $\vec{X} =_{a.s.} \mathbf{1}\vec{1}$, then $A(\vec{X}) =_{a.s.} \mathbf{1}$.*

Notice that the first condition is related to the monotonicity of usual aggregation functions and the other ones are the boundary conditions. The type of elements that can be considered is very general, the initial set S and the considered partial order $\leq$ can be adjusted for the particular scenario. An illustration of the describe construction is given in the following example.

Example 6.6 *[31] Let S be the set containing the alternatives related to the description of a speed: s_1 =slow, s_2=average, s_3=medium and s_4=fast. Intuitively, slow and fast are, respectively, the smallest and biggest speeds. However, the terms average and medium are not clearly ordered, since both refer to centrality values. Then, a reasonable partial order could be $s_1 \leq s_2$, $s_1 \leq s_3$, $s_2 \leq s_4$ and $s_3 \leq s_4$, being s_2 and s_3 incomparable. In this case, the σ-algebra $\mathscr{F}_s^n$ consists of all subsets of S^n, since any element of S can be obtained as the intersection of an upper and a lower set. Now, consider the function $A : L_S^2 \to L_S$ defined as $A(\vec{X}) = \sup_{\leq}(X_1, X_2)$, for which it is easy to check that it is well-defined.*

The latter function fulfills the three conditions given in Definition 6.5. In Table 6.1 an example of a possible distribution for $\vec{X} \in L_S^2$ and $A(\vec{X})$ is shown.

X_1 \ X_2	s_1	s_2	s_3	s_4
s_1	0.2	0	0.1	0
s_2	0.1	0	0	0
s_3	0	0.3	0	0
s_4	0.1	0.1	0	0.1

$A(\vec{X})$	s_1	s_2	s_3	s_4
P	0.2	0.1	0.1	0.6

Table 6.1: Probability distribution of a possible $\vec{X} \in L_S^2$ and its aggregation $A(\vec{X})$ [31].

6.1.1 Induced aggregations on bounded posets

As for random variables, it is possible to prove a Composition Theorem that allows one to define induced aggregations of random elements on bounded posets. Recall the notion of aggregations on bounded posets given in Definition 2.33. The following is a generalization of Theorem 3.1 in [31].

Theorem 6.7 *Let $(S, \leq, \mathbf{0}, \mathbf{1})$ be a bounded poset. Let $\hat{A} : S^n \to S$ be a measurable aggregation function with respect to the σ-algebras $\mathscr{F}_n$ and $\mathscr{F}$. Then, the function $A : L_S^n \to L_S$ defined as $A(f) = \hat{A} \circ \vec{X}$ for all $\vec{X} \in L_S^n$ is an aggregation of random elements on the bounded poset $(S, \leq, \mathbf{0}, \mathbf{1})$.*

Proof: For any $\vec{X} \in L_S^n$, $A(\vec{X}) = \hat{A} \circ \vec{X}$ is measurable as a composition of measurable functions. In addition, since $\vec{X} : \Omega \to S^n$ and $\hat{A} : S^n \to S$, it is concluded that $\hat{A} \circ \vec{X} \in L_S$, so A is well-defined.

Since $\hat{A}$ is monotone, the preimage of a measurable upper set $U \subseteq S$ is an upper subset of S^n. Consider $\vec{X}, \vec{Y} \in L_S^n$ such that $\vec{X} \leq_{st} \vec{Y}$. Then,

$$P(A(\vec{X}) \in U) = P(\vec{X} \in \hat{A}^{-1}(U)) \leq P(\vec{Y} \in \hat{A}^{-1}(U)) = P(A(\vec{Y}) \in U),$$

and therefore A is monotone.

It is clear that $\vec{X} =_{a.s.} \mathbf{0}\vec{1}$ implies that $A(\vec{X}) =_{a.s.} \mathbf{0}$ as a consequence of the boundary conditions of $\hat{A}$. The same holds for the upper boundary condition. ∎

6.1.2 Boundary coherence

A similar study as the one carried out in Sections 3.2 and 3.3 can be done for aggregations of random elements on bounded posets. In particular, notions of random and degenerate aggregations can be found in Section 3.1 in [31], following the same ideas and properties as in the case of random variables. Some coherent conditions as also defined in Section 3.2 in [31]. Although all coherence properties are not going to be disclosed here, boundary coherence is worth mentioning, since sometimes the biggest and smallest elements of the poset are especially relevant.

Definition 6.8 *[31] Let $A : L_S^n(\Omega) \to L_S(\Omega)$ be an aggregation of random elements on a bounded poset $(S, \leq, \mathbf{0}, \mathbf{1})$. Then,*

- *If for any $\vec{X} \in L_S^n(\Omega)$, $P(\vec{X} = \mathbf{0}\vec{1}) \leq P(A(\vec{X}) = \mathbf{0})$ and $P(\vec{X} = \mathbf{1}\vec{1}) \leq P(A(\vec{X}) = \mathbf{1})$, A is said to be boundary coherent,*
- *If for any $\vec{X} \in L_S^n(\Omega)$ and $\omega \in \Omega$, $\vec{X}(\omega) = \mathbf{0}\vec{1} \implies A(\vec{X})(\omega) = \mathbf{0}$ and $\vec{X}(\omega) = \mathbf{1}\vec{1} \implies A(\vec{X})(\omega) = \mathbf{1}$, then A is said to be strongly boundary coherent.*

The strongly boundary coherence can be seen as the boundaries being preserved for any element of the probability space, whereas the other one is focused on the preservation of the probability. It is straightforward to prove that strong boundary coherence implies boundary coherence. The next result states that a degenerate aggregation of random elements over a bounded poset cannot be boundary coherent.

Proposition 6.9 *[31] Let $A : L_S^n \to L_S$ be a degenerate aggregation of random elements on a bounded poset. If S has more than one element, then A is not boundary coherent.*

Proof: Since S has more than one element, then $\mathbf{0} \neq \mathbf{1}$. If $P(\vec{X} = \mathbf{0}\vec{1}) = P(\vec{X} = \mathbf{1}\vec{1}) = 0.5$ and $A(\vec{X})$ is degenerate, then it is impossible to have $P(A(\vec{X}) = \mathbf{0}) = P(A(\vec{X}) = \mathbf{1}) = 0.5$. It is concluded that A is not boundary coherent. ∎

On the other hand, any induced aggregation of random elements on a bounded poset is strongly boundary coherent.

Proposition 6.10 *[31] Let A be an induced aggregation of random elements on a bounded poset. Then, A is strongly boundary coherent.*

Proof: Let $(\Omega, \mathscr{F}, P)$ be the considered probability space. For any $\omega \in \Omega$, one has that $A(\vec{X})(\omega) = \hat{A}(\vec{X}(\omega))$. The strong boundary coherence follows by the boundary conditions of $\hat{A}$. ∎

6.2 Aggregation of random vectors

When dealing with multivariate data, sometimes one has a set of n vectors of dimension m that should be aggregated to a unique vector of the same dimension. In this direction, in recent years the theory of aggregation of vectors has been developed (see [135, 136, 265, 293, 299]).

Unlike for real numbers, there does not exist a natural total order for real vectors. The most used order is the lattice order, although this componentwise comparison is not a complete relation. In order to fix such a problem, admissible orders, which are linear extensions of the lattice order, are sometimes considered in the literature [65, 293]. The following definition of aggregation of vectors considers the one given in [293] but adapted for a possibly unbounded interval I.

Definition 6.11 *Let I be a non-empty real interval. A function $A : I^{nm} \to I^m$ is said to be an aggregation of vectors if*

1. *Is increasing,*
2. *For any $\vec{x} \in I^m$, there exists $\vec{z} \in I^{nm}$ such that $A(\vec{z}) \leq \vec{x}$,*
3. *For any $\vec{x} \in I^m$, there exists $\vec{z} \in I^{nm}$ such that $A(\vec{z}) \geq \vec{x}$.*

From a probabilistic point of view, it is possible to define a notion of aggregation of random vectors by adapting the latter definition similarly to how it has been done for random variables.

Definition 6.12 *Let (Ω,Σ,P) be a probability space and I a real non-empty interval. An aggregation function of random vectors (with respect to $\leq_{st}$) is a function $A : L_I^{nm}(\Omega) \to L_I^m(\Omega)$ which satisfies:*

1. *For any $\vec{X},\vec{Y} \in L_I^{nm}$ such that $\vec{X} \leq_{st} \vec{Y}$, $A(\vec{X}) \leq_{st} A(\vec{Y})$,*
2. *For any $\vec{X} \in L_I^m$, there exists $\vec{Z} \in L_I^{nm}$ such that $A(\vec{Z}) \leq_{st} \vec{X}$,*
3. *For any $\vec{X} \in L_I^m$, there exists $\vec{Z} \in L_I^{nm}$ such that $A(\vec{Z}) \geq_{st} \vec{X}$.*

The elements of L_I^{nm} can be seen as random matrices in which each of the rows is associated with an input vector. In particular, given n random vectors of dimension m $(\vec{X}_1,\ldots,\vec{X}_n)$, the i-th component of the j-th random vector will be denoted as X_{ji}. Aggregations of random vectors can be easily constructed considering m aggregations of random variables that are applied component by component. For instance,

Example 6.13 *The function $A : L_I^{nm} \to L_I^m$ defined as*

$$A(\vec{X}_1,\ldots,\vec{X}_n) = \left(\frac{1}{n}\sum_{i=1}^{n} X_{i1},\ldots,\frac{1}{n}\sum_{i=1}^{n} X_{im}\right),$$

is an aggregation of random vectors.

The type of aggregations of random vectors that can be decomposed into m aggregations of random variables will be called local. They have a simple structure that can be used to prove some theoretical results (see Section 6.2.2).

Definition 6.14 *An aggregation of random vectors $A : L_I^{nm} \to L_I^m$ is said to be local if there exist m aggregations of random variables $A_i : L_I^n \to L_I,\ i \in [m]$ such that:*

$$A(\vec{X}_1,\ldots,\vec{X}_n) = \left(A_1\left(\vec{X}_{11},\ldots,\vec{X}_{n1}\right),\ldots,A_m\left(\vec{X}_{m1},\ldots,\vec{X}_{m1}\right)\right),$$

for any $(\vec{X}_1,\ldots,\vec{X}_n) \in L_I^{nm}$. The aggregations of random variables $A_1,\ldots,A_m$ are said to be the components of A. In addition, if $A_1(\vec{X}) \leq_{st} \cdots \leq_{st} A_m(\vec{X})$ for any $\vec{X} \in L_I^{nm}$, A is said to be increasingly local and if $A_1 = \cdots = A_m$, A is said to be constantly local.

For instance, the aggregation of random vectors introduced in Example 6.13 is local. In particular, its components are induced by the arithmetic mean and, therefore, it is constantly local.

6.2.1 Induced aggregations of random vectors

Similar to the aggregation of random variables, a Composition Theorem analogous to Theorem 3.12 can be proved for aggregations of random vectors.

Theorem 6.15 *Let $\hat{A} : I^{nm} \to I^m$ be a measurable aggregation function of vectors. Then, the function $A : L_I^{nm} \to L_I^m$ defined as $A(\vec{X}) = \hat{A} \circ \vec{X}$ for any $\vec{X} \in L_I^{nm}$ is an aggregation of random vectors.*

Proof: Notice that, since the image of $\hat{A}$ is I^m and is measurable, A is well-defined. Monotonicity is a direct consequence of Proposition 2.106.

For the boundary conditions, consider the function $B_1 : I^{nm} \to I$ defined as $B(\vec{x}) = \min(\hat{A}(\vec{x}))$ for any $\vec{x} \in I^{nm}$. B_1 is a measurable aggregation function. Monotonicity and mesurability are a direct consequence of both A and min being increasing and measurable. For the lower boundary condition, let $x \in I$ and consider $x\vec{1} \in I^m$. Then, there exists $\vec{z} \in I^{nm}$ such that $\hat{A}(\vec{z}) \leq x\vec{1}$ and it is clear that $B(\vec{z}) = \min(\hat{A}(\vec{x})) \leq x$. For the upper boundary condition, proceed analogously. Similarly, $B_2 : I^{nm} \to I$ defined as $B_2(\vec{x}) = \max(\hat{A}(\vec{x}))$ for any $\vec{x} \in I^{nm}$ is also a measurable aggregation function.

Then, for any $\vec{X} \in L_I^m$, $\min(\vec{X}) \in L_I$. Applying Theorem 3.12 to B_2, there exists $\vec{Z} \in L_I^{nm}$ such that $B_2(\vec{Z}) = \max(\hat{A}(\vec{Z})) \leq_{st} \min(\vec{X})$. Then, it is clear that $\hat{A}(\vec{Z}) \leq_{st} \min(\vec{X})\vec{1} \leq_{st} \vec{X}$ and the lower boundary condition holds. Similarly, the upper boundary condition can be proved by using B_1. ∎

The function A is known as the aggregation of random vectors induced by the aggregation function of vectors $\hat{A}$. One may wonder if the usual stochastic order also works nicely when considering an admissible order instead of the lattice order in the aggregation of vectors. Unfortunately, the preservation of the lattice order is mandatory, and that does not hold in general for aggregations of vectors with respect to an admissible order. The following is a counterexample regarding the Xu-Yager order.

Example 6.16 *Consider the Xu-Yager order [326] $\leq_{XY}$ defined as $(x_1,x_2) \leq_{XY} (y_1,y_2)$ if and only if $x_1+x_2 < y_1+y_2$ or $x_1+x_2 = y_1+y_2$ and $x_2-x_1 \leq y_2-y_1$*

and the real interval $I = [-1,1]$. *Define the function* $A : I^4 \to I^2$ *as follows:*

$$A\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = \begin{cases} -\vec{1} & \text{if } x_{11} = x_{12} = x_{21} = x_{22} = -1, \\ \vec{1} & \text{if } x_{11} = x_{12} = x_{21} = x_{22} = 1, \\ \frac{1}{4}\begin{pmatrix} x_{11}+x_{12}+x_{21}+x_{22} \\ -(x_{11}+x_{12}+x_{21}+x_{22}) \end{pmatrix} & \text{otherwise.} \end{cases}$$

It is easy to see that A is an aggregation of vectors with respect to the Xu-Yager order. In particular, except at the extreme points, the sum of the elements of the output is always 0 and the difference is monotone with respect to the sum of the elements of the inputs. The boundary conditions are also fulfilled by definition.

Consider two random vectors of dimension 4, $\vec{X}$, *whose all its components have degenerate distribution on* 0, *and* $\vec{Y}$, *with all the components having degenerate distribution on* 0.5. *Trivially,* $\vec{X} \leq_{st} \vec{Y}$. *When A is applied,* $A(\vec{X})$ *has degenerate distribution on* $(0,0)$ *and* $A(\vec{Y})$ *has degenerate distribution on* $(0.5,-0.5)$. *Trivially,* $A(\vec{X}) \not\leq_{st} A(\vec{Y})$.

However, if there is an interest in defining induced aggregations when considering an admissible order, a possible solution could be to define a variation of the usual stochastic order as follows.

Definition 6.17 *Let* $\leq_{ad}$ *be an admissible order. Then, given two random vectors* $\vec{X}$ *and* $\vec{Y}$, *they are ordered with respect to the usual stochastic order associated with* $\leq_{ad}$, *denoted as* $\vec{X} \leq_{st-ad} \vec{Y}$, *if*

$$P(\vec{X} \in U) \leq P(\vec{Y} \in U),$$

for any upper set with respect to $\leq_{ad}$, $U \subseteq \mathbb{R}^n$.

6.2.2 Preservation of properties of random vectors

Suppose that one has n random vectors that fulfill some particular property. A natural question that arises when working with aggregations of random vectors is whenever such a property also holds for the output random vector. In this section,

some results regarding the preservation of some properties of random vectors are provided. In most of the cases, induced and local aggregations of random vectors are considered, since they have an easy structure and good properties. Starting with independence, induced and local aggregations of random vectors preserve independence of the marginals.

Proposition 6.18 *Let $A : L_I^{nm} \to L_I^m$ be an induced and local aggregation of random vectors and $\vec{X} = (\vec{X}_1, \ldots, \vec{X}_n) \in L_I^{nm}$. Let $\vec{Z}_i = \left(\vec{X}_{1i}, \ldots, \vec{X}_{ni}\right)$ with $i \in [m]$ be the random vector consisting of the i-th components of the vectors $\vec{X}_1, \ldots, \vec{X}_n$. Then, if the random vectors $\vec{Z}_1, \ldots, \vec{Z}_m$ are independent, the components of $A(\vec{X}_1, \ldots, \vec{X}_n)$ are independent.*

Proof: Since A is local and induced, $A(\vec{X}) = \left(\hat{A}_1\left(\vec{Z}_1\right), \ldots, \hat{A}_m\left(\vec{Z}_m\right)\right)$ with $A_1, \ldots, A_m : I^n \to I$ being usual aggregation functions. The result holds noticing that transformations of independent random vectors are also independent [278]. ∎

Another interesting property is exchangeability. In particular, the output random vector is exchangeable when the aggregation of random vector is constantly local and the permutations over the components of all input random vectors lead the distribution invariant. Notice that the latter condition for the input random vectors can be achieved just by assuming exchangeability of the involved random variables.

Proposition 6.19 *Let $A : L_I^{nm} \to L_I^m$ be an induced and constantly local aggregation of random vectors and $\vec{X} = (\vec{X}_1, \ldots, \vec{X}_n) \in L_I^{nm}$. Let $\vec{Z}_i = \left(\vec{X}_{1i}, \ldots, \vec{X}_{ni}\right)$ with $i \in [m]$ be the random vector consisting of the i-th components of the vectors $\vec{X}_1, \ldots, \vec{X}_n$. Then, if for any permutation $\pi : [m] \to [m]$ it holds $(\vec{Z}_1, \ldots, \vec{Z}_m) =_{st} (\vec{Z}_{\pi(1)}, \ldots, \vec{Z}_{\pi(m)})$, $A(\vec{X}_1, \ldots, \vec{X}_n)$ is exchangeable.*

Proof: Since A is induced and constantly local, there exists $\hat{A} : I^n \to I$ an aggregation function such that

$$A(\vec{X}) = \left(\hat{A}\left(\vec{Z}_1\right), \ldots, \hat{A}\left(\vec{Z}_m\right)\right).$$

Let $\pi : [m] \to [m]$ be a permutation and consider

$$\vec{Y} = \left(\hat{A}\left(\vec{Z}_{\pi(1)}\right), \ldots, \hat{A}\left(\vec{Z}_{\pi(m)}\right)\right).$$

By hypothesis, it is clear that

$$\left(\hat{A}\left(\vec{Z}_{\pi(1)}\right),\ldots,\hat{A}\left(\vec{Z}_{\pi(m)}\right)\right) =_{st} \left(\hat{A}\left(\vec{Z}_1\right),\ldots,\hat{A}\left(\vec{Z}_m\right)\right),$$

thus $A(\vec{X}) =_{st} \vec{Y}$ and $A(\vec{X})$ is exchangeable. ∎

When the components of the input random vectors are ordered in the usual stochastic order, applying an increasingly local aggregation of random vectors preserves such an ordering.

Proposition 6.20 *Let $A : L_I^{nm} \to L_I^m$ be an increasingly local aggregation of random vectors and $\vec{X} = (\vec{X}_1,\ldots,\vec{X}_n) \in L_I^{nm}$. Let $\vec{Z}_i = \left(\vec{X}_{1i},\ldots,\vec{X}_{ni}\right)$ with $i \in [m]$ be the random vector consisting of the i-th components of the vectors $\vec{X}_1,\ldots,\vec{X}_n$. Then, if $\vec{Z}_1 \leq_{st} \cdots \leq_{st} \vec{Z}_m$, then $A(\vec{X})_1 \leq_{st} \cdots \leq_{st} A(\vec{X})_m$.*

Proof: From the definition of increasingly local and the monotonicity of aggregations of random variables, $A_1(\vec{Z}_1) \leq_{st} \cdots \leq_{st} A_m(\vec{Z}_m)$, where $A_1,\ldots,A_n$ are the components of A. ∎

Notice that if $\vec{Z}_{\pi(1)} \leq_{st} \cdots \leq_{st} \vec{Z}_{\pi(m)}$ for a permutation $\pi : [m] \to [m]$, the latter result holds when the components of A are ordered as $A_{\pi(1)}(\vec{X}) \leq_{st} \cdots \leq_{st} A_{\pi(m)}(\vec{X})$ for any $\vec{X} \in L_I^{nm}$.

Finally, the existence of covariance matrices is relevant for many applications. For instance, concepts such as efficiency (see Definition 2.122) need the associated variances to be finite. Fortunately, if the covariance matrix of a random vector exists, so does the covariance matrix of the result of applying an aggregation of random vectors that is local and induced by internal aggregation functions.

Proposition 6.21 *Let $A : L_I^{nm} \to L_I^m$ be an induced and local aggregation of random vectors such that its components $A_1,\ldots,A_m$ are internal. Then, for any $\vec{X} \in L_I^{nm}$ for which its covariance matrix exists, then the covariance matrix of $A(\vec{X})$ exists.*

Proof: Denote $\vec{Z}_1,\ldots,\vec{Z}_m$ as in the latter proofs. Since A has induced and internal components $A_1,\ldots,A_m$

$$\min\left(\vec{Z}_j\right) \leq A_j\left(\vec{Z}_j\right) = A(\vec{X})_j \leq \max\left(\vec{Z}_j\right).$$

Therefore, the following inequality is reached,

$$\left(A_j\left(\vec{Z}_j\right)\right)^2 \leq \left(\max\left(\vec{Z}_j\right)\right)^2 + \left(\min\left(\vec{Z}_j\right)\right)^2.$$

Divide $\mathbb{R}^m$ into the subsets $U_i = \{\vec{x} \in \mathbb{R}^n : \max(\vec{x}) = x_i \text{ and } x_i > x_k,\ k \in [i-1]\}$, with $i \in [n]$. Similarly, consider also the subsets $L_i = \{\vec{x} \in \mathbb{R}^m : \min(\vec{x}) = x_i \text{ and } x_i < x_k,\ k \in [i-1]\}$, with $i \in [n]$. These subsets are measurable, see [23]. Let $f(\vec{z})$ be the density function of $\vec{Z}_j$. The second moment of $A(\vec{X})_j$ can be bounded as follows

$$\begin{aligned} E\left[\left(A_j\left(\vec{Z}_j\right)\right)^2\right] &= \int_{\mathbb{R}^n} \left(A_j\left(\vec{z}\right)\right)^2 f\left(\vec{z}\right) d\vec{z} \leq \\ &\leq \int_{\mathbb{R}^n} \left(\left(\max\left(\vec{z}\right)\right)^2 + \left(\min\left(\vec{z}\right)\right)^2\right) f(\vec{z}) d\vec{z} = \\ &= \sum_{i=1}^{n} \left(\int_{U_i} z_i^2 f(\vec{z}) d\vec{z} + \int_{L_i} z_i^2 f(\vec{z}) d\vec{z}\right) \leq 2 \sum_{i=1}^{n} E\left[\vec{X}_{ji}^2\right] \end{aligned}$$

The latter bound is well defined and finite since $\vec{X}$ has a finite covariance matrix, thus its components have a finite second moment. Therefore, the components of $A(\vec{X})$ have finite second moments, which implies finite first moments by Jensen inequality (see [171]) and also finite variance. Since all the variances are finite, the same holds for the covariances. It is concluded that $A(\vec{X})$ has a finite covariance matrix. ∎

6.3 Aggregation of stochastic processes

In the last section, the aggregation of random variables is extended to aggregate n finite collections of random variables, that is, random vectors. A natural step in this regard is to consider now possibly infinite collections of random variables, stochastic processes.

Stochastic processes can be seen as random functions. Unfortunately, it is not possible to find in the literature works focused on the aggregation of functions. The closest notion could be the one introduced in Appendix A in [149], in which the aggregation of an infinite number of inputs is considered. However, the result of such an aggregation is a real number, not another function. Therefore, the first step

to introduce an aggregation of stochastic processes is to define the aggregation of functions.

Given a real interval I, consider a function from an arbitrary set T to I. The set of this type of functions will be denoted as I^T. It will be said that $f, g : T \to I$ are ordered, $f \leq g$, if $f(t) \leq g(t)$ for any $t \in T$, which is the natural partial order for functions. When n-dimensional functions of the form $f : T \to I^n$ are considered, the associated set will be denoted as $I^T \times \cdots \times I^T = I^{nT}$. Then, the extension from aggregations of vectors to aggregations of functions is immediate.

Definition 6.22 *Let I be a real interval and T be an arbitrary set. A function $A : I^{nT} \to I^T$ is said to be an aggregation of functions if it satisfies:*

1. *For any $f, g \in I^{nT}$ such that $f \leq g$, $A(f) \leq A(g)$,*
2. *For any $f \in I^T$, there exists $g \in I^{nT}$ such that $A(g) \leq f$,*
3. *For any $f \in I^T$, there exists $g \in I^{nT}$ such that $A(g) \geq f$.*

Although this definition does not require any condition for the set T, in the following it will be supposed that T is an interval of the real line or an infinite subset of consecutive integers, since many stochastic processes are indexed over a set related to time and some of their properties (see Definitions 2.89 and 2.90), need these types of index sets. Although stochastic processes are defined as a collection of random variables $(X_t, t \in T)$, in the following they will be denoted just as X_t to ease the notation. In addition, if a random vector is associated with each of the indices, it will be denoted as $\vec{X}_t$ and its components will be denoted as $X_{i,t}$.

However, when working with an infinite index set T, there is the possibility of having non-bounded functions. As will be seen in Section 6.3.1, this could be a problem. In this direction, bounded aggregations of functions can be defined just by replacing the set I^{nT} for $I_b^{nT} = \{f \in I^{nT} \mid f \text{ is bounded } \}$ and I^T for $I_b^T = \{f \in I^T \mid f \text{ is bounded } \}$ in Definition 6.22.

Moving to the probabilistic setting, the associated sets for stochastic processes can be defined by following the same idea.

Definition 6.23 *Let (Ω,Σ,P) be a probability space, I a real interval and T a set. Then, the set of all stochastic processes defined over (Ω,Σ,P) with sample paths in I^{nT} is defined as:*

$$L_I^{nT}(\Omega)=\left\{\vec{X}_t : T\times\Omega\to I^n \mid \vec{X}_t \text{ is measurable } \forall t\in T\right\}.$$

Similarly to other cases, it will be denoted $L_I^{nT}(\Omega)$ just by L_I^{nT} and L_I^{1T} by L_I^T. If the interval I is unbounded, it is possible to have stochastic processes in L_I^T with non-bounded sample paths. This would be a problem in order to prove boundary conditions of induced aggregations of stochastic processes. In this direction, the following set is defined.

$$B_I^{nT}(\Omega)=\left\{\vec{X}_t\in L_I^{nT}(\Omega) \;\middle|\; \sup_{t\in T}X_{t,i},\inf_{t\in T}X_{t,i}\in L_I(\Omega) \text{ for any } i\in[n]\right\}.$$

By requiring $\sup_{t\in T}X_{t,i}\in L_I$ and $\inf_{t\in T}X_{t,i}\in L_I$, it is imposed that the supremum and infimum should exist and be random variables. For the existence of the supremum and the infimum, it is enough to have that the processes are bounded. For the measurability, it is enough to impose that the processes are separable (see [25]). Moreover, any stochastic process has a separable modification (another stochastic process with the same finite-dimensional distributions that is separable) [90, 92]. Similarly as in the previous case, consider the notation shortcuts B_I^{nT} and B_I^T. Notice that if I is bounded, then B_I^{nT} and L_I^{nT} are essentially the same.

The elements of L_I^{nT} can be seen as vector valued stochastic processes. In particular, given $\vec{X}_t\in L_I^{nT}$, the stochastic process associated with the i-th component will be denoted as $X_{i,t}$. With all the elements already settled, the notion of aggregation of stochastic processes can be defined.

Definition 6.24 *Let (Ω,Σ,P) be a probability space and I a real non-empty interval. An aggregation function of stochastic processes is a function $A:L_I^{nT}(\Omega)\to L_I^T(\Omega)$ which satisfies:*

1. *For any $\vec{X}_t,\vec{Y}_t\in L_I^{nT}$ such that $\vec{X}_t\leq_{st}\vec{Y}_t$, $A(\vec{X}_t)\leq_{st}A(\vec{Y}_t)$,*
2. *For any $X_t\in L_I^T$, there exists $\vec{X}_t\in L_I^{nT}$ such that $A(\vec{X}_t)\leq_{st}X_t$,*
3. *For any $X_t\in L_I^T$, there exists $\vec{X}_t\in L_I^{nT}$ such that $A(\vec{X}_t)\geq_{st}X_t$.*

If the sets B_I^{nT} and B_I^T are considered instead of L_I^{nT} and L_I^{nT} in the latter definition, it is said that A is an aggregation of bounded stochastic processes. Aggregations of stochastic processes can be easily constructed by considering a family of aggregations of random variables that is applied to each of the indices. For instance,

Example 6.25 *The function $A : L_I^{nT} \to L_I^T$ defined as*

$$A(\vec{X}_t) = \frac{1}{n}\sum_{i=1}^{n} X_{i,t},$$

is an aggregation of stochastic processes.

In direction, a notion of locality, as already defined for aggregation of random vectors, can be considered for aggregations of stochastic processes.

Definition 6.26 *An aggregation of stochastic processes $A : L_I^{nT} \to L_I^T$ is said to be local if there exists a family of aggregations of random variables $(A_t : L_I^n \to L_I,\ t \in T)$ such that:*

$$A(\vec{X})_t = A_t\left(\vec{X}_t\right).$$

The aggregations of random variables $(A_t : L_I^n \to L_I,\ t \in T)$ are said to be the components of A. In addition, if $A_{t_1} = A_{t_2}$ for any $t_1, t_2 \in T$, A is said to be constantly local.

For instance, the aggregation of stochastic processes given in Example 6.25 is local. In particular, its components are induced by the arithmetic mean and, therefore, it is constantly local. The latter definition can be extended straightforwardly to aggregations of bounded stochastic processes.

6.3.1 Induced aggregations of stochastic processes

As stated before, the set L_I^T, when I is unbounded, includes unbounded stochastic processes. This is a problem when trying to prove a Composition Theorem for stochastic processes, since in the proof the supremum and infimum of the involved stochastic processes are used when dealing with the boundary conditions. However, when working with B_I^{nT}, everything works as intended. The following is a generalization of Theorem 3.12 and Theorem 6.15.

Theorem 6.27 *Let $\hat{A} : I_b^{n^T} \to I_b^T$ be a bounded measurable aggregation function of functions. Then, the function $A : B_I^{n^T} \to B_I^T$ defined as $A(\vec{X}_t) = \hat{A} \circ \vec{X}_t = \hat{A}(\vec{X})$ is an aggregation of bounded stochastic processes.*

Proof: Notice that, since the image of $\hat{A}$ is I_b^T and is measurable, A is well-defined.

For monotonicity, consider $\vec{X}_t \in B_I^{n^T}$ and $\vec{Y}_t, \in B_I^{n^T}$ two stochastic processes such that $\vec{X}_t \leq_{st} \vec{Y}_t$. Then, using Theorem 2.113, there exist $\hat{\vec{X}}_t$ and $\hat{\vec{Y}}_t$ such that $\hat{\vec{X}}_t =_{st} \vec{X}_t$, $\hat{\vec{Y}}_t =_{st} \vec{Y}_t$ and $\hat{\vec{X}}_t =_{a.s.} \hat{\vec{Y}}_t$. Since $\hat{A}$ is increasing, it is clear that $A(\hat{\vec{X}}_t) \leq_{a.s.} A(\hat{\vec{Y}}_t)$. Then, since $\hat{A}(\hat{\vec{X}}_t) =_{st} A(\vec{X}_t)$ and $\hat{A}(\hat{\vec{Y}}_t) \leq_{st} A(\vec{Y}_t)$, it is concluded that $A(\vec{X}_t) \leq_{st} A(\vec{Y}_t)$.

For the boundary conditions, consider the function $C_1 : I \to I$ fulfilling $C_1(x) = \inf_{\vec{t} \in T^n} \hat{A}(f(\vec{t}))$ with f being the constant function such that $f(\vec{t}) = x$ for any $\vec{t} \in T^n$. It is clear that f is bounded. C_1 is well-defined since $\hat{A}(f)$ is bounded. Moreover, according to the definition of B_I^T, if X is a random variable, then $C_1 \circ X$ is also a random variable.

C_1 is an aggregation function (of dimension 1). The monotonicity is a direct consequence of the monotonicity of $\hat{A}$. For the lower boundary condition, consider $x \in I$. Then, the function defined as $f(\vec{t}) = x$ for any $\vec{t} \in T^n$ is bounded. Then, by the boundary conditions of $\hat{A}$, there exists a bounded $g \in I^{n^T}$ such that $\hat{A}(g) \leq f$. Let $y = \inf_{\vec{t} \in T^n} g(\vec{t})$, which exists since g is bounded. Then, it is clear that $C_1(y) \leq \inf_{t \in T} \hat{A}(g) \leq x$. For the upper boundary condition, proceed analogously. Similarly, $C_2 : I \to I$ defined as $C_2(x) = \sup_{\vec{t} \in T^n} \hat{A}(f(\vec{t}))$ with f being the constant function such that $f(\vec{t}) = x$ for any $\vec{t} \in T^n$ is an aggregation function.

Let $X_t \in B_I^T$. By the definition of B_I^T, $\inf_{t \in T} X_t \in L_I$. Applying Theorem 3.12 to C_2, there exists $Y \in L_I$ such that $C_2 \circ Y \leq_{st} \inf_{t \in T} X_t$. Consider the stochastic process $\vec{Y}_t \in B_I^{n^T}$ defined as $Y_{i,t} = Y$ for any $t \in T$ and $i \in [n]$. Then, $C_2(Y) = \sup_{t \in T} \hat{A}(\vec{Y}_t) \leq_{st} \inf_{t \in T} X_t$. Then, it is clear that $\hat{A}(\vec{Y}_t) \leq_{st} X_t$. Similarly, the upper boundary condition can be proved by using C_1. ∎

6.3.2 Preservation of properties of stochastic processes

Similarly to the study already considered in Section 6.2.2, there are many properties of stochastic processes that are of interest. In some cases, one does not want to lose

these properties when applying an aggregation of stochastic processes. In the following, sufficient conditions for the preservation of several properties are provided, mainly related to being induced and local.

As a first step, it is easy to show that the Markov property is not preserved. This is a consequence of the fact that for the Markov property to be preserved, bijective functions should be considered. To the best knowledge of the author, for instance, there does not exist any increasing bijection between $[0,1]^n$ and $[0,1]$, thus no aggregation function can fulfill this role. However, a weaker result can be proved.

Proposition 6.28 *Let $A : L_{I^T}^n \to L_{I^T}$ be an induced and local aggregation of stochastic processes and let $\vec{X}_t \in L_I^{nT}$. Then, if $\vec{X}_t$ fulfills the Markov property, then $A(\vec{X}_t)_{t_0}$ and $A(\vec{X}_t)_{t_2}$ are conditionally independent given $\vec{X}_{t_1}$ for any $t_0, t_1, t_2 \in T$ such that $t_0 \leq t_1 \leq t_2$.*

Proof: Since $\vec{X}_t$ is Markovian, it holds that $\vec{X}_{t_0}$ and $\vec{X}_{t_2}$ are conditionally independent given $\vec{X}_{t_1}$ for any $t_0, t_1, t_2 \in T$ such that $t_0 \leq t_1 \leq t_2$. Since A is local and induced, there exists two measurable aggregation functions A_{t_0} and A_{t_2} such that $A_{t_0}(\vec{X}_{t_0}) = A(\vec{X}_t)_{t_0}$ and $A_{t_2}(\vec{X}_{t_2}) = A(\vec{X}_t)_{t_2}$. Since (conditional) independence is preserved when composing a random variable and a measurable function (see [278]) it holds that $A(\vec{X}_t)_{t_0}$ and $A(\vec{X}_t)_{t_2}$ are conditionally independent given $\vec{X}_{t_1}$. ∎

Secondly, the property of being stationary, i.e. having a distribution that is invariant with respect to translations, is preserved when applying an aggregation of stochastic processes that is also invariant in this regard.

Proposition 6.29 *Let $A : L_I^{nT} \to L_I^T$ be an induced and constantly local aggregation of stochastic processes and let $\vec{X}_t \in L_I^{nT}$. Then, if $\vec{X}_t$ is stationary, then $A(\vec{X}_t)$ is stationary.*

Proof: Consider two collections of random vectors $(\vec{X}_{t_1}, \ldots, \vec{X}_{t_s})$ and $(\vec{X}_{t_1+h}, \ldots, \vec{X}_{t_s+h})$. Since $\vec{X}_t$ is stationary, both have the same distribution. Applying that A is constantly local and induced,

$$A(\vec{X})_{t_1}, \ldots, A(\vec{X})_{t_s} = \hat{A} \circ \vec{X}_{t_1}, \ldots, \hat{A} \circ \vec{X}_{t_s},$$
$$A(\vec{X})_{t_1+h}, \ldots, A(\vec{X})_{t_s+h} = \hat{A} \circ \vec{X}_{t_1+h}, \ldots, \hat{A} \circ \vec{X}_{t_s+h},$$

where $\hat{A}$ is an aggregation function. Trivially, the latter random vectors have the same distribution, thus it is concluded that $A(\vec{X}_t)$ is stationary. ∎

Ergodicity is a relevant property, since it allows one to estimate the constant expectation of the stochastic process just by using one sample path instead of many independent observations. Linearity is crucial for this property, so it is preserved when applying weighted arithmetic means.

Proposition 6.30 *Let $A : L_I^{nT} \to L_I^T$ be a constantly local aggregation of stochastic processes induced by a weighted arithmetic mean and let $\vec{X}_t \in L_I^{nT}$. If $\vec{X}_t$ is ergodic, then $A(\vec{X}_t)$ is ergodic.*

Proof: Suppose that the mean vector of $\vec{X}_t$ is $\vec{\mu}_t$, which is constant for any $t \in T$ since the process is ergodic. Given a weighting vector $\vec{w}$ for the weighted arithmetic mean, $E[A(\vec{X}_t)] = \sum_{i=1}^n w_i \mu_i$.

In the case that $T = [0, \infty)$, compute the time average estimation:

$$\lim_{L\to\infty} \frac{1}{L} \int_0^L A(\vec{X})_t dt = \lim_{L\to\infty} \frac{1}{L} \int_0^L \sum_{i=1}^n w_i X_{i,t} dt = \sum_{i=1}^n w_i \lim_{L\to\infty} \frac{1}{L} \int_0^L X_{i,t} dt =_{a.s} \sum_{i=1}^n w_i \mu_i,$$

where the last step is achieved by using the ergodicity of $\vec{X}_t$. Therefore, $A(\vec{X}_t)$ is also ergodic. If T is discrete, the proof is achieved analogously. ∎

Similarly, being a martingale is defined in terms of an expectation, which has good properties with respect to linear functions. Therefore, weighted arithmetic means applied to martingales return martingales.

Proposition 6.31 *Let $A : L_I^{nT} \to L_I^T$ be a constantly local aggregation of stochastic processes induced by a weighted arithmetic mean and let $\vec{X}_t \in L_I^{nT}$. If $\vec{X}_t$ is a martingale, then $A(\vec{X}_t)$ is a martingale.*

Proof: Consider $t_0, t \in T$ such that $t \geq t_0$. Suppose that $A(\vec{X})_{t_0}$ equals λ. Denote as $g(\vec{x})$ the conditional density function of $\vec{X}_{t_0}$ given $A(\vec{X})_{t_0} = \lambda$. The mean of $A(\vec{X})_t$ given $A(\vec{X})_{t_0} = \lambda$ can be computed as follows:

$$\int E[A(\vec{X})_t \mid \vec{X}_{t_0} = \vec{x}] g(\vec{x}) d\vec{x}.$$

Then, compute $E[A(\vec{X})_t \mid \vec{X}_{t_0} = \vec{x}]$ for a fixed $\vec{x} \in \mathbb{R}^n$ such that $\sum_{i=1}^n w_i x_i = \lambda$. Applying the fact that $\vec{X}_t$ is a martingale,

$$E[A(\vec{X})_t \mid \vec{X}_{t_0} = \vec{x}] = E\left[\sum_{i=1}^n w_i X_{i,t} \;\middle|\; \vec{X}_{t_0} = \vec{x}\right] =$$
$$= \sum_{i=1}^n w_i E[X_{i,t} \mid \vec{X}_{t_0} = \vec{x}] = \sum_{i=1}^n w_i x_i = \lambda.$$

Then, the initial integral can be computed as follows:

$$\int E[A(\vec{X})_t \mid \vec{X}_{t_0} = \vec{x}] g(\vec{x}) d\vec{x} = \int \lambda\, g(\vec{x}) d\vec{x} = \lambda.$$

The same holds when considering the conditional distribution given, in addition, the values of the vectors $\vec{X}_{t_1}, \ldots, \vec{X}_{t_k}$ such that $t_i \leq t_0$ for any $i \in [k]$ and $k \in \mathbb{N}$. It is concluded that $A(\vec{X})$ is a martingale. The case of discrete distributions or discrete index set can be proven analogously. ∎

In addition, the properties of being a submartingale or supermartingale are preserved when, respectively, the maximum and minimum are applied to each index.

Proposition 6.32 *Let $A : L_I^{nT} \to L_I^T$ be a constantly local aggregation of stochastic processes induced by the maximum (minimum) and let $\vec{X}_t \in L_I^{nT}$. If $\vec{X}_t$ is a submartingale (supermartingale), then $A(\vec{X}_t)$ is a submartingale (supermartingale).*

Proof: Consider $t_0, t \in T$ such that $t \geq t_0$. Let $(\Omega, \mathscr{F}, P)$ be the probability space. Consider the disjoint decomposition of Ω given by $\Omega = \cup_{i=1}^n C_i$ with:

$$C_i = \left\{ w \in \Omega : \max(\vec{X}_{t_0}) = X_{i,t_0},\ \max(\vec{X}_{t_0}) \neq X_{j,t_0},\ j < i \right\}.$$

Of course, C_i is measurable for any $i \in [n]$. Then, the inequality for the mean can be computed for every subset C_i, using that $\vec{X}_t$ is a submartingale:

$$E\left[\max(\vec{X}_t) \;\middle|\; \max(\vec{X}_{t_0}) = \lambda, C_i\right] = E\left[\max(\vec{X}_t) \;\middle|\; X_{i,t_0} = \lambda\right] \geq$$
$$\geq E\left[\vec{X}_{i,t} \;\middle|\; X_{i,t_0} = \lambda\right] \geq \lambda.$$

Finally, the total expectation is computed:

$$E\left[\max(\vec{X}_t) \;\middle|\; \max(\vec{X}_{t_0}) = \lambda\right] = \sum_{i=1}^{n} E\left[\max(\vec{X}_t) \;\middle|\; \max(\vec{X}_{t_0}) = \lambda, C_i\right] P(C_i) \geq$$
$$\geq \sum_{i=1}^{n} \lambda P(C_i) = \lambda.$$

The same holds when considering the conditional distribution given, in addition, the values of the vectors $\vec{X}_{t_1}, \ldots, \vec{X}_{t_k}$ such that $t_i \leq t_0$ for any $i \in [k]$ and $k \in \mathbb{N}$. The proof for the minimum is analogous. ∎

Finally, it is shown that the reflection principle, fulfilled, for instance, by the Brownian motion [168], is preserved by the minimum.

Proposition 6.33 *Let $A : L_I^{nT} \to L_I^T$ be a constantly local aggregation of stochastic processes induced by the minimum and let $\vec{X}_t \in L_I^{nT}$. If the components of $\vec{X}_t$ are independent and fulfill the reflection principle, then $A(\vec{X}_t)$ fulfills the reflection principle.*

Proof: Compute the probability associated with the reflection principle. Notice that, if the aggregation is the minimum, the aggregated value is greater than or equal to a value a if and only if all the inputs are also greater than or equal to a. Also notice the independence of the components of $\vec{X}_t$, which allows one to decompose the joint probability as the product of the marginal probabilities.

$$P\left(\sup_{0\leq t\leq s} A(\vec{X})_t \geq a\vec{1}\right) = \prod_{i=1}^{n} P\left(\sup_{0\leq t\leq s} X_{i,t} \geq a\right) =$$
$$= \prod_{i=1}^{n} 2P\left(X_{i,s} \geq a\right) = 2P\left(A(\vec{X})_s \geq a\vec{1}\right).$$

∎

6.4 Aggregation of random sets

The aggregation of random sets, random elements that take values in a set of subsets, is disclosed in this section. In the following, non-empty closed subsets of $\mathbb{R}^n$ will

be considered.

The main obstacle to defining a notion of aggregation of random sets is that there does not exist a suitable location stochastic order for random sets already defined. Some approaches in the literature consider dispersion stochastic ordering for random intervals [83], dispersion stochastic orders defined by the Aumann expectation [127] or location stochastic relations (but not orders), between p-boxes [242].

Therefore, the first step in the definition of aggregations of random sets will be to establish a good definition of a stochastic order for random sets. This construction, as well as its main properties, will be disclosed in the next section. The proper definition of aggregation of random sets, as well as a Composition Theorem, will be introduced in Section 6.4.2.

6.4.1 The usual stochastic order for random sets

This section is devoted to the definition of the usual stochastic order for closed, non-empty and Effros measurable random sets. As a preliminary step, one should consider an order for closed sets. Recall that F^n denotes the set of closed subsets of $\mathbb{R}^n$.

Definition 6.34 *[34] Let $C_1, C_2 \in F^n$ be two closed sets. Then, if for any $\vec{x}_1 \in C_1$ and $\vec{x}_2 \in C_2$ it holds $\vec{x}_1 \leq \vec{x}_2$ or $C_1 = C_2$, it is said that C_1 is smaller than or equal to C_2 and it is denoted as $C_1 \leq C_2$.*

Although another order could be chosen, this particular choice is interesting since, when extended to random sets, it has almost all the desirable properties that one expects. It will be denoted as $C_1 < C_2$ when $C_1 \leq C_2$ but $C_1 \neq C_2$. As a first step, it should be proved that the latter relation is a partial order.

Proposition 6.35 *[34] The relation $\leq$ defined in Definition 6.34 is a partial order.*

Proof: Reflexivity and transitivity are immediate. For the antisymmetry, suppose that $C_1 \neq C_2$, $C_1 \leq C_2$ and $C_2 \leq C_1$. Then, if $\vec{x}_1 \in C_1$ and $\vec{x}_2 \in C_2$, one has that $\vec{x}_1 \leq \vec{x}_2$ and $\vec{x}_2 \leq \vec{x}_1$. Then, C_1 and C_2 consist of just one point and $C_1 = C_2$, which is a contradiction. ∎

The usual stochastic order for random variables and vectors is defined, among other alternative definitions, in terms of inequalities of probabilities over measurable upper sets (see Theorem 2.102). In the next definition, the notion is extended by considering upper sets with respect to the partial order of Definition 6.34.

Definition 6.36 *[34] Let S_X and S_Y be two random sets of dimension n. Then, if for any measurable upper set (with respect to $\leq$) U it is hold that*

$$P(S_X \in U) \leq P(S_Y \in U),$$

it is said that S_X is smaller than S_Y with respect to the usual stochastic order for random sets and it is denoted as $S_X \leq_{st} S_Y$.

In general, it is not easy to check if an upper set on F^n is measurable or not. The most simple examples are of the form $\{C \in F^n \mid C \cap U \neq \emptyset\}$ and $\{C \in F^n \mid C \subseteq U\}$, which coincide with the sets of closed subsets that, respectively, intersect or are contained in an upper set U of $\mathbb{R}^n$.

Proposition 6.37 *[34] Let S_X and S_Y be two random sets such that $S_X \leq_{st} S_Y$. Then, for any upper set $U \subseteq \mathbb{R}^n$,*

(1) $P(S_X \cap U \neq \emptyset) \leq P(S_Y \cap U \neq \emptyset)$,

(2) $P(S_X \subseteq U) \leq P(S_Y \subseteq U)$.

Proof: It is enough to prove that $\{C \in F^n \mid C \cap U \neq \emptyset\}$ and $\{C \in F^n \mid C \subseteq U\}$ are upper sets of F^n. For the first case, consider $C_1 \in \{C \in F^n \mid C \cap U \neq \emptyset\}$ and $C_2 \in F^n$ such that $C_1 \leq C_2$. Let $\vec{x}_1 \in C_1 \cap U$. Then, since C_2 is not empty, there exists $\vec{x}_2 \in C_2$ such that $\vec{x}_1 \leq \vec{x}_2$. Since U is an upper set of $\mathbb{R}^n$, then $\vec{x}_2 \in U$, so $\vec{x}_2 \in C_2 \cap U$. It is concluded that $C_2 \cap U \neq \emptyset$.

For the second case, let $C_1 \in \{C \in F^n \mid C \subseteq U\}$ and $C_2 \in F^n$ be such that $C_1 \leq C_2$. Consider any $\vec{x}_2 \in C_2$. Then there exists $\vec{x}_1 \in C_1$ such that $\vec{x}_1 \leq \vec{x}_2$. Using that $C_1 \subseteq U$, one has $\vec{x}_1 \in U$, so $\vec{x}_2 \in U$. It is concluded that $C_2 \subseteq U$. ∎

The interpretation of the result above is immediate. If $S_X \leq_{st} S_Y$, S_X intersects upper sets of $\mathbb{R}^n$ with less probability than S_Y. Similarly, S_X has a smaller probability of being contained in an upper set of $\mathbb{R}^n$ than S_Y.

Before studying in more detail the usual stochastic order for random sets, it remains to prove that the relation defined is, indeed, a stochastic order in the sense of Definition 2.100. Reflexivity and transitivity can be proved straightforwardly. However, antisymmetry is far from being straightforward. In this regard, it is necessary to prove that, if $S_X =_{st} S_Y$, then they have the same distribution. The proof will be constructed in subsequent results. As a first step, consider n-dimensional open cubes.

Lemma 6.38 *[34] Let S_X and S_Y be two random sets such that $S_X =_{st} S_Y$ and $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^n$ such that $\vec{x}_1 < \vec{x}_2$. Then, $P(S_X \cap C \neq \emptyset) = P(S_Y \cap C \neq \emptyset)$ with $C = \{\vec{x} \in \mathbb{R}^n \mid \vec{x}_1 < \vec{x} < \vec{x}_2\}$.*

Proof: Define $L = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} < \vec{x}_2\}$ and $U = \{\vec{x} \in \mathbb{R}^n \mid \vec{x}_1 < \vec{x}\}$. Notice that U is an upper set of $\mathbb{R}^n$ and L is a lower set of $\mathbb{R}^n$. Decompose the probability for S_X as follows,

$$P(S_X \cap C \neq \emptyset) = P(S_X \cap (U \cap L) \neq \emptyset) = P(S_X \cap (U \backslash (\bar{L} \cap U)) \neq \emptyset) = \\ = P(S_X \cap U \neq \emptyset) - P(S_X \cap U \neq \emptyset, S_X \cap U \subseteq \bar{L}).$$

Similarly, one has $P(S_Y \cap C \neq \emptyset) = P(S_Y \cap U \neq \emptyset) - P(S_Y \cap U \neq \emptyset, S_Y \cap U \subseteq \bar{L})$. Using Proposition 6.37, $P(S_X \cap U \neq \emptyset) = P(S_Y \cap U \neq \emptyset)$. It remains to prove that $\{C \in F^n \mid C \cap U \neq \emptyset, C \cap U \subseteq \bar{L}\}$ is an upper set in F^n.

Let $C_1 \in \{C \in F^n \mid C \cap U \neq \emptyset, C \cap U \subseteq \bar{L}\}$ and $C_2 \in F^n$ be such that $C_1 \leq C_2$. As proven in Proposition 6.37, it holds that $C_2 \cap U \neq \emptyset$.

Consider $\vec{x}_1 \in C_1 \cap U$ and $\vec{x}_2 \in C_2 \cap U$. Notice that, since $C_1 \leq C_2$, then $C_1 \cap U \leq C_2 \cap U$ and $\vec{x}_1 \leq \vec{x}_2$. Then, using the same procedure as in (2) in Proposition 6.37 by noticing that $\bar{L}$ is an upper set, it holds that $\vec{x}_2 \in \bar{L}$ and it is concluded that $C_2 \cap U \subseteq \bar{L}$.

Then, $P(S_X \cap U \neq \emptyset, S_X \cap U \subseteq \bar{L}) = P(S_Y \cap U \neq \emptyset, S_Y \cap U \subseteq \bar{L})$ and it is concluded that $P(S_X \cap C \neq \emptyset) = P(S_Y \cap C \neq \emptyset)$. ∎

The next step is to generalize the result to the intersection, simultaneously, with a finite number of n-dimensional open cubes. The result holds by expressing such probabilities as a finite linear combination of the probabilities of the form of the latter result.

Lemma 6.39 *[34] Let $C_1,\ldots,C_m$ be sets defined as Lemma 6.38 with $m \in \mathbb{N}$. If $S_X =_{st} S_Y$, then it holds that*

$$P(S_X \cap C_1 \neq \emptyset,\ldots,S_X \cap C_m \neq \emptyset) = P(S_Y \cap C_1 \neq \emptyset,\ldots,S_Y \cap C_m \neq \emptyset).$$

Proof: Decompose each C_i with $i \in [m]$ as $C_i = U_i \cap L_i$ as in Lemma 6.38. The following identity will be proved:

$$P(S_X \cap C_1 \neq \emptyset,\ldots,S_X \cap C_m \neq \emptyset) = \sum_{\vec{x}\in\mathbb{B}(m)} (-1)^{\sum_{i=1}^n x_i} P(A_1^{x_1},\ldots,A_m^{x_m}),$$

where A_i^0 is the event $S_X \cap U_i \neq \emptyset$, A_i^1 is the intersection of the events $S_X \cap U_i \neq \emptyset$ and $S_X \cap U_i \subseteq \bar{L}_i$ and $\mathbb{B}(m) = \{\vec{x} \in \mathbb{R}^m \mid x_i \in \{0,1\}\}$ is the boolean cube of dimension m. The case $m = 1$ has already been proven in Proposition 6.37. Suppose that it is true for $m-1$. Then,

$$\begin{aligned}
&P(S_X \cap C_1 \neq \emptyset,\ldots,S_X \cap C_m \neq \emptyset) = \\
&= \sum_{\vec{x}\in\mathbb{B}(m-1)} (-1)^{\sum_{i=1}^{m-1} x_i} P(A_1^{x_1},\ldots,A_{m-1}^{x_{m-1}},S_X \cap C_m \neq \emptyset) = \\
&= \sum_{\vec{x}\in\mathbb{B}(m-1)} (-1)^{\sum_{i=1}^{m-1} x_i} \big[P(A_1^{x_1},\ldots,A_{m-1}^{x_{m-1}},S_X \cap U_m \neq \emptyset) - \\
&\quad -P(A_1^{x_1},\ldots,A_{m-1}^{x_{m-1}},S_X \cap U_m \neq \emptyset, S_X \cap U_m \subseteq \bar{L}_m)\big] = \\
&\qquad = \sum_{\vec{x}\in\mathbb{B}(m)} (-1)^{\sum_{i=1}^{m} x_i} P(A_1^{x_1},\ldots,A_m^{x_m}).
\end{aligned}$$

The result follows by noticing that the sets A_i^0 and A_i^1 are upper sets for any $i \in [m]$ and that the intersection of upper sets is an upper set. ∎

The last step before the main result is to consider probabilities related to the intersection with at least one of these cubes instead to all of them.

Lemma 6.40 *[34] Let $C_1,\ldots,C_m$ be sets defined as in Lemma 6.38. If $S_X =_{st} S_Y$, then it holds that*

$$P(S_X \cap (\cup_{i=1}^n C_i) \neq \emptyset) = P(S_Y \cap (\cup_{i=1}^n C_i) \neq \emptyset).$$

Proof: It will be proved that:

$$P(S_X \cap (\cup_{i=1}^{m} C_i) \neq \emptyset) = \sum_{\vec{x} \subseteq [m]} (-1)^{\#\vec{x}} P(S_X \cap C_{x_1} \neq \emptyset, \ldots, S_X \cap C_{x_{\#\vec{x}}} \neq \emptyset),$$

where $\#\vec{x}$ denotes the number of elements of $\vec{x}$.

For $m = 1$, it holds since the right side of the latter equation is just $P(S_X \cap C_1 \neq \emptyset)$. Suppose that it is true for $m-1$. Then,

$$\begin{aligned}
P(S_X \cap (\cup_{i=1}^{m} C_i) \neq \emptyset) &= P(S_X \cap \left(\left(\cup_{i=1}^{m-1} C_i\right) \cup C_m\right) \neq \emptyset) = \\
&= P(S_X \cap \left(\cup_{i=1}^{m-1} C_i\right) \neq \emptyset) + P(S_X \cap C_m \neq \emptyset) - \\
-P(S_X \cap \left(\cup_{i=1}^{m-1} C_i\right) \neq \emptyset, S_X \cap C_m \neq \emptyset) &= P(S_X \cap C_m \neq \emptyset) + \\
+ \sum_{\vec{x} \subseteq [m-1]} (-1)^{\#\vec{x}} \Big[P(S_X \cap C_{x_1} \neq \emptyset, \ldots, S_X \cap C_{x_{\#\vec{x}}} \neq \emptyset) - \\
-(-1)^{\#\vec{x}} P(S_X \cap C_{x_1} \neq \emptyset, \ldots, S_X \cap C_{x_{\#\vec{x}}} \neq \emptyset, S_X \cap C_m \neq \emptyset) \Big] = \\
= \sum_{\vec{x} \subseteq [m]} (-1)^{\#\vec{x}} P(S_X \cap C_{x_1} \neq \emptyset, \ldots, S_X \cap C_{x_{\#\vec{x}}} \neq \emptyset).
\end{aligned}$$

Finally, the identity $P(S_X \cap \left(\cup_{i=1}^{m} C_i\right) \neq \emptyset) = P(S_Y \cap \left(\cup_{i=1}^{m} C_i\right) \neq \emptyset)$ holds since all the summands of the latter equation are of the form of Lemma 6.39, thus have the same value for S_X and for S_Y. ∎

The latter results lead to the main one, in which it is proved that S_X and S_Y have the same distribution if and only if they hold $S_X =_{st} S_Y$. In particular, open covers of compact sets consisting on open cubes are considered, proving the result by using the Choquet Theorem.

Theorem 6.41 *[34] Let S_X and S_Y be two random sets. Then, $S_X =_{st} S_Y$ if and only if S_X and S_Y have the same distribution.*

Proof: Trivially, if S_X and S_Y have the same distribution, then $S_X =_{st} S_Y$. For the other implication, let K be a compact set on $\mathbb{R}^n$. Consider $R_k = \cup_{\vec{p} \in K} B_M(\vec{p}, \frac{1}{k})$ with

$$B_M\left(\vec{p}, \frac{1}{k}\right) = \left\{ \vec{x} \in \mathbb{R}^n \;\middle|\; \max_{i \in [n]} |x_i - p_i| < \frac{1}{k} \right\},$$

being the ball with center $\vec{p} \in K$ and radius $\frac{1}{n}$ with respect to the Manhattan distance.

Notice that $\left(B_M(\vec{p}, \frac{1}{k}), \vec{p} \in K\right)$ is, for any $k \in \mathbb{N}$, an open cover of K. Since K is compact, there exists a finite open subcover, i.e. $\left(B_M(\vec{p}, \frac{1}{n}), \vec{p} \in K_k\right)$ with K_k finite and such that $K \subseteq \cup_{\vec{p} \in K_k} B_M(\vec{p}, \frac{1}{n})$.

Consider $R'_k = \cup_{\vec{p} \in K_k} B_M(\vec{p}, \frac{1}{k})$. In the following, the convergence $\lim_{k\to\infty} R_k = K$ will be proved. It is clear that $K \subset R_k$ for any $k \in \mathbb{N}$. In addition, let $\vec{x} \in \lim_{n\to\infty} R_k$. Then, for any $k \in \mathbb{N}$, there exists $\vec{p} \in K$ such that $\max_{i\in[n]} |x_i - p_i| < \frac{1}{k}$. Therefore, $d(\vec{x}, K) = 0$. Since K is compact, it is closed and therefore $\vec{x} \in K$.

Then, $\lim_{k\to\infty} R_k \subseteq K$. It is concluded that $\lim_{k\to\infty} R_k = K$. Moreover, since $R_{k+1} \subseteq R_k$ for any $k \in \mathbb{R}$, one has that for any random set S

$$\lim_{k\to\infty} P(S \cap R_k \neq \emptyset) = P\left(S \cap \left(\lim_{k\to\infty} R_k\right) \neq \emptyset\right) = P(S \cap K \neq \emptyset).$$

Notice that, since $K \subseteq R'_k \subseteq R_k$ for any $k \in \mathbb{N}$, it is clear that:

$$P(S \cap K \neq \emptyset) \leq \lim_{k\to\infty} P(S \cap R'_k \neq \emptyset) \leq \lim_{k\to\infty} P(S \cap R_k \neq \emptyset) = P(S \cap K \neq \emptyset),$$

thus $P(S \cap K \neq \emptyset) = \lim_{k\to\infty} P(S \cap R'_k \neq \emptyset)$. Noticing that $R'_k = \cup_{\vec{p} \in K_k} B_M(\vec{p}, \frac{1}{k})$ is a set of the form described in Proposition 6.40 for any $k \in \mathbb{N}$, it is concluded that $P(S_X \cap K \neq \emptyset) = P(S_Y \cap K \neq \emptyset)$. The result holds by using Theorem 2.92. ∎

Therefore, the usual stochastic order for random sets is a stochastic order as defined in Definition 2.100. Similarly as with the usual stochastic order of random vectors, having $P(S_X \leq S_Y) = 1$ ensures $S_X \leq_{st} S_Y$.

Proposition 6.42 *[34] Let S_X and S_Y be two random sets defined in the same probability space such that $P(S_X \leq S_Y) = 1$ (denoted as $S_X \leq_{a.s.} S_Y$). Then, $S_X \leq_{st} S_Y$.*

Proof: Let U be an upper set of F^n and let $C \subseteq \Omega$ be the measurable subset of the probability space such that, for any $\omega \in C$, $S_X(\omega) \leq S_Y(\omega)$. Then, if $S_X(\omega) \in U$, $S_Y(\omega) \in U$ for any $\omega \in C$. Since $P(C) = 1$, then $P(S_X \in U) \leq P(S_Y \in U)$ and the result holds. ∎

Using the latter result, it is easy to construct pairs of random sets that are ordered with respect to the usual stochastic order. The next result involves a Poisson process.

Example 6.43 *[34] Let $(\Phi_t, t \in \mathbb{R}^+)$ be a homogeneous Poisson process (see [188]) with rate λ. Then, consider $n \in \mathbb{N}$ and the random sets $[X_{n-1}, X_n]$ and $[X_n, X_{n+1}]$, where X_n denotes the arrival time of Φ_t at the value $n \in \mathbb{N}$. It holds that $P([X_{n-1}, X_n] \leq [X_n, X_{n+1}]) = 1$, so $[X_{n-1}, X_n] \leq_{st} [X_n, X_{n+1}]$.*

6.4.1.1 Closure properties

In the following results, the preservation of the usual stochastic order for random sets is studied. Some of the here-presented results are generalizations of the well-known cases for random random vectors, which can be found in Theorem 6.B.16. of [295]. The first case that is going to be provided is the closure with respect to mixtures.

Proposition 6.44 *[34] Let S_X, S_Y be two random sets and let $\vec{\Theta}$ be a random vector such that $[S_X \mid \vec{\Theta} = \vec{\theta}] \leq_{st} [S_Y \mid \vec{\Theta} = \vec{\theta}]$ for any $\vec{\theta} \in S(\vec{\Theta})$. Then, $S_X \leq_{st} S_Y$.*

Proof: Let U be any upper set on F^n. Then,

$$P(S_X \in U) = E[P(S_X \in U \mid \vec{\Theta})] \leq E[P(S_Y \in U \mid \vec{\Theta})] = P(S_Y \in U).$$

■

Another relevant property of the usual stochastic order for random vectors is that it is preserved when an increasing function is applied (see Proposition 2.106). Notice that the notion of increasing function should be linked to an order, which in this case is the one previously introduced in Definition 6.34.

Definition 6.45 *[34] A function $h : F^n \to F^m$ is said to be increasing if for any $C_1, C_2 \in F^n$ such that $C_1 \leq C_2$, it holds that $h(C_1) \leq h(C_2)$.*

A sufficient condition to have an increasing function in terms of the latter definition is to apply a closed (i.e. the image of closed sets is closed) and increasing function of real numbers to closed sets.

Proposition 6.46 *[34] Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be an increasing and closed function. Then, the function $f' : F^n \to F^m$ defined as $f'(C) = \{f(\vec{x}) \mid \vec{x} \in C\}$ is increasing.*

Proof: Since f is closed, the image of a closed set is closed, thus f' is well-defined. If $C_1 = C_2$, it is clear that $f'(C_1) = f'(C_2)$. If $C_1 \leq C_2$, then $\vec{x}_1 \leq \vec{x}_2$ for any $\vec{x}_1 \in C_1$ and $\vec{x}_2 \in C_2$. Since f is increasing, then $f(\vec{x}_1) \leq f(\vec{x}_2)$ for any $\vec{x}_1 \in C_1$ and $\vec{x}_2 \in C_2$. Then, $f'(C_1) \leq f'(C_2)$. ■

In the next result, the closure of the usual stochastic order for random sets with respect to the application of increasing functions is stated.

Theorem 6.47 *[34] Let S_X and S_Y be two random sets such that $S_X \leq_{st} S_Y$ and $f : F^n \to F^m$ be a measurable increasing function. Then, $f(S_X) \leq_{st} f(S_Y)$.*

Proof: Let U be any upper set of F^m. Then,

$$P(f(S_X) \in U) = P(S_X \in f^{-1}(U)) \leq P(S_Y \in f^{-1}(U)) = P(f(S_Y) \in U),$$

where it has been used that the preimage of an upper set by an increasing function is also an upper set. ■

In the following, it is explored whenever the usual stochastic order for random vectors is preserved for the typical set operations. In this direction, not all the properties that one could expect to be fulfilled actually hold.

The first one is the Cartesian product. Unfortunately, it is easy to find counterexamples for this case. For instance, $\{0\} \times [0,1] \not\leq \{1\} \times [0,1]$, since $(0,0.5) \in \{0\} \times [0,1]$ and $(1,0.2) \in \{0\} \times [0,1]$, even it is clear that $\{0\} \leq \{1\}$ and $[0,1] \leq [0,1]$. The best one can do is to prove the following result.

Proposition 6.48 *[34] Let S_X^1, S_X^2, S_Y^1 and S_Y^2 be random sets such that $S_X^1 <_{a.s.} S_Y^1$ and $S_X^2 <_{a.s.} S_Y^2$. Then, $S_X^1 \times S_X^2 <_{a.s.} S_Y^1 \times S_Y^2$.*

Proof: There exists a subset $B \subseteq \Omega$ in the probability space such that, for any $\omega \in B$, one has that $(\vec{x}_1, \vec{x}_2) \leq (\vec{x}_1, \vec{x}_2)$ for any $\vec{x}_1 \in S_X^1(\omega)$, $\vec{x}_2 \in S_X^2(\omega)$, $\vec{y}_1 \in S_Y^1(\omega)$ and $\vec{y}_2 \in S_Y^2(\omega)$. Then, it is clear that $(\vec{x}_1, \vec{x}_2) \leq (\vec{y}_1, \vec{y}_2)$ and $S_X^1 \times S_X^2 <_{a.s.} S_Y^1 \times S_Y^2$. ■

The intersection has better behavior. It is important to remark that, since non-empty sets have been considered, the intersection of the random sets should be, at

least, non-empty with probability one. In particular, having two random sets S_1 and S_2 of the same dimension such that $P(S_1 \cap S_2 \neq \emptyset) = 1$, one can define:

$$S_X = \begin{cases} S_1 \cap S_2 & \text{if } S_1 \cap S_2 \neq \emptyset, \\ \{0\} & \text{otherwise.} \end{cases}$$

The value $\{0\}$ could be any other closed subset, since it would take this value with probability 0 and, for probability computations, is negligible. Then, it is clear that S_X does not take empty values and fulfills $S_X =_{st} S_1 \cap S_2$. It turns out that the intersection preserves the usual stochastic order for random sets.

Proposition 6.49 *[34] Let S_X^1, S_X^2, S_Y^1 and S_Y^2 be random sets of the same dimension such that, $S_X^1 \cap S_X^2 \neq_{a.s.} \emptyset$, $S_Y^1 \cap S_Y^2 \neq_{a.s.} \emptyset$ and $S_X^1 \times S_X^2 \leq_{st} S_Y^1 \times S_Y^2$. Then, $S_X^1 \cap S_X^2 \leq_{st} S_Y^1 \cap S_Y^2$.*

Proof: Define $h : F^{2n} \to F^n$ as $h(C_1, C_2) = C_1 \cap C_2$. It should be proved that h is increasing. Consider $C_1, C_2, C_1', C_2' \in F^n$ such that $C_1 \leq C_1'$ and $C_2 \leq C_2'$. Consider $\vec{x} \in C_1 \cap C_2$ and $\vec{y} \in C_1' \cap C_2'$. If $C_1 < C_1'$, since $\vec{x} \in C_1$ and $\vec{y} \in C_1'$, $\vec{x} \leq \vec{y}$. If $C_2 < C_2'$, one similarly has that $\vec{x} \leq \vec{y}$. Finally, if $C_1 = C_1'$ and $C_2 = C_2'$, then $C_1 \cap C_2 = C_1' \cap C_2'$. Then, the result is a consequence of Theorem 6.47. ∎

The last operation that will be studied is the union. Similarly to the Cartesian product, one can find counterexamples for this case. For instance, $\{0\} \cup \{2\} \not\leq \{1\} \cup \{3\}$ while $\{0\} \leq \{1\}$ and $\{2\} \leq \{3\}$. However, weaker properties can be proved when working with independent random sets.

Proposition 6.50 *[34] Let S_X^1, S_X^2, S_Y^1 and S_Y^2 be independent random sets of the same dimension such that $S_X^1 \leq_{st} S_Y^1$ and $S_X^2 \leq_{st} S_Y^2$. Then, for any upper set U of $\mathbb{R}^n$,*

1. $P((S_X^1 \cup S_X^2) \cap U \neq \emptyset) \leq P((S_Y^1 \cup S_Y^2) \cap U \neq \emptyset)$,
2. $P((S_X^1 \cup S_X^2) \subseteq U \neq \emptyset) \leq P((S_Y^1 \cup S_Y^2) \subseteq U)$.

Proof: For the first property,

$$P((S_X^1 \cup S_X^2) \cap U \neq \emptyset) = P(S_X^1 \cap U \neq \emptyset) + P(S_X^2 \cap U \neq \emptyset) -$$
$$-P(S_X^1 \cap U \neq \emptyset, S_X^2 \cap U \neq \emptyset) = P(S_X^1 \cap U \neq \emptyset) + P(S_X^2 \cap U \neq \emptyset) -$$
$$-P(S_X^1 \cap U \neq \emptyset) P(S_X^2 \cap U \neq \emptyset) \leq P(S_Y^1 \cap U \neq \emptyset) + P(S_Y^2 \cap U \neq \emptyset) -$$
$$-P(S_Y^1 \cap U \neq \emptyset, S_Y^2 \cap U \neq \emptyset) = P((S_Y^1 \cup S_Y^2) \cap U \neq \emptyset).$$

For the second one,

$$P((S_X^1 \cup S_X^2) \subseteq U) = P(S_X^1 \subseteq U, S_X^1 \subseteq U) = P(S_X^1 \subseteq U) P(S_X^1 \subseteq U) \leq$$
$$\leq P(S_Y^1 \subseteq U) P(S_Y^1 \subseteq U) = P((S_Y^1 \cup S_Y^2) \subseteq U).$$

■

6.4.1.2 Selections and expectation

The usual stochastic order has a characterization in terms of the expectation of increasing functions (see Theorem 2.102). In this section, some results related to the selection expectation and the usual stochastic order for random sets are provided. As a first step, consider the comparison of a finite number of selection operators.

Proposition 6.51 *[34] Let S_X and S_Y be two random sets such that $S_X \leq_{st} S_Y$ and let $f_1, \ldots, f_k : F^n \to \mathbb{R}^n$ be selection operators. Then,*

$$(f_1(S_X), \ldots, f_k(S_X)) \leq_{st} (f_1(S_Y), \ldots, f_k(S_Y)).$$

Proof: It is enough to prove that for any $C_1, C_2 \in F^n$ such that $C_1 \leq C_2$, $(f_1(C_1), \ldots, f_k(C_1)) \leq (f_1(C_2), \ldots, f_k(C_2))$. If $C_1 = C_2$, it is clear that $(f_1(C_1), \ldots, f_k(C_1)) = (f_1(C_2), \ldots, f_k(C_2))$. If $C_1 < C_2$, since $f_i(C_1) \in C_1$ and $f_i(C_2) \in C_2$ for any $i \in [k]$ and for any $\vec{x}_1 \in C_1$ and $\vec{x}_2 \in C_2$ it holds $\vec{x}_1 \leq \vec{x}_2$, then one has $(f_1(C_1), \ldots, f_k(C_1)) \leq (f_1(C_2), \ldots, f_k(C_2))$. The result is reached by using Theorem 6.47. ■

As a direct consequence, if a countable sequence of selectors $(f_i, i \in \mathbb{N})$ is chosen, $(f_i(S_X), i \in \mathbb{N}) \leq_{st} (f_i(S_Y), i \in \mathbb{N})$, where here $\leq_{st}$ stands for the usual stochastic

order for stochastic processes (see Definition 2.112). Recall the definition of integrable random sets and integrable selections given right after Definition 2.96. In the next result, it is proved that integrable selections of ordered random sets are ordered.

Theorem 6.52 *[34] Let S_X and S_Y be integrable random sets such that $S_X \leq_{st} S_Y$. Then,*

- *For any integrable selection of S_X, $\vec{X}$, there exists an integrable selection of S_Y, $\vec{Y}$, such that $\vec{X} \leq_{st} \vec{Y}$,*
- *For any integrable selection of S_Y, $\vec{Y}$, there exists an integrable selection of S_X, $\vec{X}$, such that $\vec{X} \leq_{st} \vec{Y}$.*

Proof: By Proposition 2.95, there exists a family of selection operators $(f_i, i \in \mathbb{N})$ such that $S_X = cl\{f_i(S_X), i \in \mathbb{N}\}$ and $S_Y = cl\{f_i(S_Y), i \in \mathbb{N}\}$. In addition, as stated before, one has that $(f_i(S_X), i \in \mathbb{N}) \leq_{st} (f_i(S_Y), i \in \mathbb{N})$ as stochastic processes. Let $\vec{X}_0$ be an integrable selection of S_X and $\vec{Y}_0$ an integrable selection of S_Y. Then, define

$$\vec{X}'_{i,j} = \mathbf{1}_{||\vec{X}_n||_2 \in [j-1,j)]} f_i(S_X) + \mathbf{1}_{||\vec{X}_n||_2 \notin [j-1,j)]} \vec{X}_0,$$
$$\vec{Y}'_{i,j} = \mathbf{1}_{||\vec{Y}_n||_2 \in [j-1,j)]} f_i(S_Y) + \mathbf{1}_{||\vec{Y}_n||_2 \notin [j-1,j)]} \vec{Y}_0.$$

The elements of the sequences $(\vec{X}'_{i,j},\ i,j \in \mathbb{N})$ and $(\vec{Y}'_{i,j},\ i,j \in \mathbb{N})$ are integrable. Moreover, $S_X = cl\{\vec{X}'_{i,j},\ i,j \in \mathbb{N}\}$ and $S_Y = cl\{\vec{Y}'_{i,j},\ i,j \in \mathbb{N}\}$ (see the proof of Proposition 1.2. of Chapter 2 in [239]). Since both sequences are numerable, rearrange the elements so $\{\vec{X}''_i,\ i \in \mathbb{N}\} = \{\vec{X}'_{i,j},\ i,j \in \mathbb{N}\}$ and $\{\vec{Y}''_i,\ i \in \mathbb{N}\} = \{\vec{Y}'_{i,j},\ i,j \in \mathbb{N}\}$.

Let $\vec{X} \in S_X$ be a selection of S_X. Then, it is also a selection of $cl\{\vec{X}''_i,\ i \in \mathbb{N}\}$. Denote as $\vec{Z}_n$ the random variable $\sum_{j=1}^{m(n)} \mathbf{1}_{A_j} \vec{X}''_j$ of Theorem 2.98 when $p = 1$ and $\varepsilon = \frac{1}{n}$. Then, it is clear that:

$$E\left[||\vec{X} - \vec{Z}_n||_2\right] \to 0,$$

and, therefore, since 2-norm convergence implies convergence in distribution [222], $\vec{X} =_{st} \lim_{n\to\infty} \vec{Z}_n$.

Now, denote as $\vec{W}_n$ the random vector $\sum_{j=1}^{m(n)} \mathbf{1}_{A_j} \vec{Y}''_j$, where the subsets $A_1, \ldots, A_l$ are the same as in $\vec{Z}_n$. Then, define $\vec{Y} = \lim_{n\to\infty} \vec{W}_n$, which is a random variable

since it is the limit of a sequence of random variables. For each $\omega \in \Omega$, one has that $\vec{Y}(\omega) = \lim_{n\to\infty} \sum_{j=1}^{m(n)} \mathbf{1}_{A_j} \vec{Y}_j''(\omega)$. That is, $\vec{Y}(\omega)$ is a limit point of $\{Y_i''(\omega), i \in \mathbb{N}\}$ and, therefore, $\vec{Y}(\omega) \in cl\{Y_i''(\omega), i \in \mathbb{N}\}$, so is a selection of S_Y. Finally, since $\vec{X}_i'' \leq_{st} \vec{Y}_i''$ for any $i \in \mathbb{N}$, applying Propositions 2.104 and 2.107 it holds $\vec{X} \leq_{st} \vec{Y}$.

The other property can be proved analogously by considering $-S_X$ and $-S_Y$. ■

Also in the context of selections, one might ask whether the selection expectation (see Definition 2.99) of two ordered random sets is also ordered. Unfortunately, the answer is negative, since the sum of two sets is not increasing in the sense of Definition 6.45.

Example 6.53 *[34] Let S_X be a random set that takes the values $[0,1]$ and $[1,2]$ with probability 0.5. Similarly, consider S_Y the random set taking the values $[1,2]$ and $[1,3]$ with probability 0.5.*

Trivially, these sets fulfill $S_X \leq_{st} S_Y$, since they can be seen as a mixture of the cases $[1,2] = [1,2]$ and $[0,1] \leq [1,3]$, see Proposition 6.44. However, the selection expectations are $E[S_X] = [0.5, 1.5]$ and $E[S_Y] = [1, 2.5]$ and one has that $[0.5, 1.5] \not\leq [1, 2.5]$.

However, a weaker result follows directly from Theorem 6.52.

Corollary 6.54 *[34] Let S_X and S_Y be two integrable random sets such that $S_X \leq_{st} S_Y$. Then,*

- *For any $\vec{x} \in E[S_X]$, there exists $\vec{y} \in E[S_Y]$ such that $\vec{x} \leq \vec{y}$,*
- *For any $\vec{y} \in E[S_Y]$, there exists $\vec{x} \in E[S_X]$ such that $\vec{x} \leq \vec{y}$.*

6.4.1.3 Relation with other stochastic orders

Every notion generalized from random vectors to random sets should, if possible, be equivalent to the initial one if the values of random set are singletons. This is the case of the usual stochastic order for random sets.

Proposition 6.55 *[34] Let $\vec{X}$ and $\vec{Y}$ be two random vectors. Then, $\vec{X} \leq_{st} \vec{Y}$ if and only if $\{\vec{X}\} \leq_{st} \{\vec{Y}\}$.*

Proof: For the first implication, suppose that $\vec{X} \leq_{st} \vec{Y}$ and let U be any upper set of F^n. Consider $U' = \cup_{A \subseteq U, |A|=1} A$, that is, the set consisting of the points associated with the singletons sets of U. Then, $P(\{\vec{X}\} \in U) = P(\vec{X} \in U')$, $P(\{\vec{Y}\} \in U) = P(\vec{Y} \in U')$ and therefore,

$$P(\{\vec{X}\} \in U) = P(\vec{X} \in U') \leq P(\vec{Y} \in U') = P(\{\vec{Y}\} \in U),$$

and it is concluded that $\{\vec{X}\} \leq_{st} \{\vec{Y}\}$.

For the second one, suppose that $\{\vec{X}\} \leq_{st} \{\vec{Y}\}$. Let U' be any upper set of $\mathbb{R}^n$. Then, by using (2) in Proposition 6.37,

$$P(\vec{X} \in U) = P(\{\vec{X}\} \subseteq U) \leq P(\{\vec{Y}\} \subseteq U) = P(\vec{Y} \in U),$$

and it is concluded that $\vec{X} \leq_{st} \vec{Y}$. ∎

As introduced in Definition 2.114, some alternatives for the stochastic ordering of random intervals, in the context of imprecise probabilities, were proposed in [242]. The relation between these stochastic relations and the proposed definition is explored in the following result.

Proposition 6.56 *[34] Let $I_X = [X_1, X_2]$ and $I_Y = [Y_1, Y_2]$ be two random intervals. Then,*

$$I_X \leq_{FSD_1} I_Y \implies I_X \leq_{st} I_Y \implies I_X \leq_{FSD_2} I_Y \text{ and } I_X \leq_{FSD_5} I_Y.$$

Proof: For the first implication, since $X_2 \leq_{st} Y_1$, $X_1 \leq_{a.s.} X_2$ and $Y_1 \leq_{a.s.} Y_2$, it holds that $(X_1, X_2) \leq_{a.s.} (X_2, X_2) \leq_{st} (Y_1, Y_1) \leq_{a.s.} (Y_1, Y_2)$, thus $(X_1, X_2, X_2) \leq_{st} (Y_1, Y_1, Y_2)$.

Applying Theorem 2.102, consider the two random vectors $(\hat{X}_1, \hat{X}_2, \hat{X}_2)$ and $(\hat{Y}_1, \hat{Y}_1, \hat{Y}_2)$ fulfilling that $(\hat{X}_1, \hat{X}_2, \hat{X}_2) =_{st} (X_1, X_2, X_2)$, $(\hat{Y}_1, \hat{Y}_1, \hat{Y}_2) =_{st} (Y_1, Y_1, Y_2)$ and $(\hat{X}_1, \hat{X}_2, \hat{X}_2) \leq_{a.s.} (\hat{Y}_1, \hat{Y}_1, \hat{Y}_2)$.

Then, $[\hat{X}_1, \hat{X}_2] \leq_{a.s} [\hat{Y}_1, \hat{Y}_2]$, $[X_1, X_2] =_{st} [\hat{X}_1, \hat{X}_2] \leq_{a.s} [\hat{Y}_1, \hat{Y}_2] =_{st} [X_1, X_2]$ and it is concluded that $I_X \leq_{st} I_Y$.

For the second implication, since $I_X \leq_{st} I_Y$, applying Proposition 6.37 one has that $P(I_X \subseteq (a, \infty)) \leq P(I_Y \subseteq (a, \infty))$ for any $a \in \mathbb{R}$. Therefore, since $P(I_X \subseteq (a, \infty)) = P(X_1 > a)$ and $P(I_Y \subseteq (a, \infty)) = P(Y_1 > a)$, $P(X_1 > a) \leq P(Y_1 > a)$. This

implies that the distribution function of X_1 is pointwise greater than the distribution function of Y_1 and it is concluded that $X_1 \leq_{st} Y_1$ and $I_X \leq_{FSD_2} I_Y$.

Similarly, applying again Proposition 6.37, one has that $P(I_X \cap [a,\infty)) \leq P(I_Y \cap [a,\infty) \neq \emptyset)$ for any $a \in \mathbb{R}$. Then, since $P(I_X \cap [a,\infty) \neq \emptyset) = P(X_2 \geq a)$ and $P(I_Y \cap [a,\infty) \neq \emptyset) = P(Y_2 \geq a)$, $P(X_2 \geq a) \leq P(Y_2 \geq a)$. Analogously as in the previous case, it is concluded that $I_X \leq_{FSD_5} I_Y$. ∎

Notice that other stochastic relations $(\leq_{FSD_3}, \leq_{FSD_4}, \leq_{FSD_6})$ are also defined in [242], all of them being implied by $\leq_{FSD_2}$ or $\leq_{FSD_5}$. However, none of them are equal to $\leq_{st}$ when restricted to random intervals, since they are not transitive, reflexive and antisymmetric with respect to having the same distribution.

6.4.2 Aggregation functions of random sets

After the detailed study of the usual stochastic order for random sets, a notion of aggregation of these random structures can be defined. First of all, a modification of the sets L_I^n, defined for random vectors, should be considered. In particular, given a probability space $(\Omega, \mathscr{F}, P)$, define the set $\mathscr{L}_I^n(\Omega)$ as

$$\mathscr{L}_I^n(\Omega) = \{S_X : \Omega \to F^n \mid S_X \text{ is non-empty, Effros measurable, closed and } P(S_X \subseteq I) = 1\}.$$

As usual for other structures, they will be denoted for simplicity as $\mathscr{L}_I^n$ and $\mathscr{L}_I^1 = \mathscr{L}_I$. Another point that should be addressed are the boundary conditions. Notice that for unbounded random sets, it is not possible to define proper boundary conditions. For instance, there does not exist any $S \in F^n$ such that $S \leq (-\infty, 0]$ or $S \geq [0,\infty)$, see Definition 6.34. However, it still makes sense to impose boundary conditions for bounded sets.

Finally, monotonicity is defined using $\leq_{st}$ in Definition 6.36. With all these elements, a notion of aggregation of random sets can be introduced as follows.

Definition 6.57 *Let (Ω, Σ, P) be a probability space and I a real non-empty interval. An aggregation function of random sets (with respect to $\leq_{st}$) is a function $A : \mathscr{L}_I^n(\Omega) \to \mathscr{L}_I(\Omega)$ which satisfies:*

1. *For any $S_X, S_Y \in \mathscr{L}_I^n$ such that $S_X \leq_{st} S_Y$, $A(S_X) \leq_{st} A(S_Y)$,*
2. *For any almost surely lower bounded $S_1 \in \mathscr{L}_I$, there exists $S_2 \in \mathscr{L}_I^n$ such that $A(S_2) \leq_{st} S_1$,*
3. *For any almost surely upper bounded $S_1 \in \mathscr{L}_I$, there exists $S_2 \in \mathscr{L}_I^n$ such that $A(S_2) \geq_{st} S_1$.*

When working with random vectors, the latter conditions are equivalent to the ones of Definition 3.6. The first equivalence is a consequence of Proposition 6.55, while the second and the third hold noticing that a set consisting of just one point is always bounded.

As seen in the last examples of aggregation of random structures, it would be adequate to have a Composition Theorem adapted to random sets that allows one to construct aggregation of random sets by applying a non-random function to the random sets. Recall the definition of monotonicity introduced in Definition 6.45.

Theorem 6.58 *Let $\hat{A} : F^n \to F$ be an increasing measurable function such that*

1. *For any lower bounded $S \in F$, there exists a lower bounded $S_2 \in F^n$ verifying $\hat{A}(S_2) \leq S_1$,*
2. *For any upper bounded $S \in F$, there exists an upper bounded $S_2 \in F^n$ verifying $\hat{A}(S_2) \geq S_1$.*

Then, the function $A : \mathscr{L}_I^n \to \mathscr{L}_I$ defined as $A(S_X) = \hat{A} \circ S_X = \hat{A}(\vec{X})$ is an aggregation of random sets.

Proof: Notice that, for any $S_X \in \mathscr{L}_I^n$, $S_X : \Omega \to F^n$ and, since $\hat{A} : F^n \to F$, then $\hat{A}(\vec{X}) : \Omega \to F$. The measurability of $\hat{A}(S_X)$ is a consequence of the measurability of both S_X and $\hat{A}$. Therefore, $\hat{A}(\vec{X}) \in \mathscr{L}_I$ and A is well-defined.

Monotonicity is a direct consequence of Theorem 6.47.

For the boundary conditions, consider the function $C_1 : I \to I$ such that $C_1(x) = \inf \hat{A}(\{x\} \times \cdots \times \{x\})$. C_1 is well-defined since $\hat{A}(\{x\} \times \cdots \times \{x\})$ is lower bounded. Moreover, using Theorem 2.93, if X is a random variable, then $C_1 \circ X$ is also a random variable.

C_1 is an aggregation function (of dimension 1). The monotonicity is a direct consequence of the monotonicity of $\hat{A}$. For the lower boundary condition, consider $x \in I$. Then, by the properties of $\hat{A}$, there exists a lower bounded $S \in F^n$ such that $\hat{A}(S) \leq \{x\}$. Let $y = \inf_{\vec{x} \in S} \min(\vec{x})$, which exists since S is lower bounded. Then, it is clear that $C_1(y) \leq \inf \hat{A}(S) \leq x$. For the upper boundary condition, proceed analogously. Similarly, $C_2 : I \to I$ defined as $C_2(x) = \sup \hat{A}(\{x\} \times \cdots \times \{x\})$ is an aggregation function.

Let $S_X \in \mathscr{L}_I$. Applying Theorem 2.93, $\inf S_X \in L_I$. Applying Theorem 3.12 to C_2, there exists $Y \in L_I$ such that $C_2 \circ Y \leq_{st} \inf S_X$. Then, $C_2(Y) = \sup \hat{A}(\{Y\} \times \cdots \times \{Y\}) \leq_{st} \inf S_X$. Then, it is clear that $\hat{A}(\{Y\} \times \cdots \times \{Y\}) \leq_{st} S_X$. Similarly, the upper boundary condition can be proved by using C_1. ∎

6.5 Aggregation of additional structures

In the following, some brief comments are provided for other random structures including random intervals, random graphs, random positive semi-definite matrices and fuzzy random variables.

6.5.1 Aggregation of random intervals

Closed intervals can be seen as vectors of dimension two fulfilling that the first element is always smaller or equal to the second one. In the literature, great attention has been paid to the aggregation of intervals [113, 134, 263]. These aggregations are used when the data is not a precise number but an interval containing the real value, which is unknown.

The aggregation of random intervals can be studied by considering two different approaches, a modification of aggregations of random vectors or a modification of aggregations of random sets. In the first case, it is enough to restrict the study to bivariate random vectors (X_1, X_2) such that $X_1 \leq_{a.s.} X_2$. Then a similar result as in Theorem 6.15 can be stated. The monotonicity is trivially fulfilled, while for the boundary conditions notice that the random vector $\vec{Z}$ in the proof of Theorem 6.15 can be replaced by $\max(\vec{Z})\vec{1}$, which, if $n = 2$, can be seen as a degenerate random

interval. In the second case, random intervals can be seen as a particular case of random closed sets. When more than one random interval is considered, they can be seen as the Cartesian product of such intervals, which is a random closed set of dimension n. Clearly, these random sets are contained in $\mathscr{L}_I^n(\Omega)$, so Theorem 6.58 can be used.

Among these two alternatives, the first one considers a weaker stochastic order for the intervals, since, if $I_X = [X_1, X_2]$ and $I_Y = [Y_1, Y_2]$ are two random intervals, $X_1 \leq_{st} Y_1$ and $X_2 \leq_{st} Y_2$ are required. In the second one, the usual stochastic order for sets is stronger, since $X_2 \leq_{st} Y_1$ is sometimes required. The monotonicity of the first can be applied to more vectors of random intervals, but implies a weaker ordering of the result of the aggregation. The second one states a stronger comparison between the result of the aggregation, but it can be applied to fewer cases. It has to be said that in the literature the usual approach is to consider intervals as particular cases of two dimensional vectors, not as sets. Therefore, the first approach should be considered in this regard.

The most remarkable example of random intervals is confidence intervals in estimation [7, 302]. Given a random sample, an interval of possible values of a parameter is given. Of course, the interval inherits the randomness of the random sample. In this scenario, aggregation procedures could be useful. If two confidence intervals for the same parameter have been constructed from two different samples of the same populations, it is reasonable to combine both confidence intervals to obtain a new one that uses the information of two samples, as done in [223]. This procedure can be seen as an aggregation of the initial ones.

6.5.2 Aggregation of random graphs

The aggregation of graphs can be considered when applying a voting rule, the graph being a preference relation [13], when searching for consensus between different clustering of a dataset [128] or in the study of social networks [13]. This concept has been introduced in [122], without considering, in general, any monotonicity condition. The particular case of monotonic aggregation of (directed) graphs will be introduced, in which the partial order is the inclusion order of the edges.

Definition 6.59 *[300] Let $V = [m]$ be a set of nodes and let $\mathscr{G}$ be the set of all*

graphs of the form (V,E), with $E \subseteq V \times V$. Then, the inclusion order is defined as:

$$(V,E_1) \leq (V,E_2) \iff E_1 \subseteq E_2.$$

In this regard, $(\mathscr{G},\leq,(V,\emptyset),(V,V\times V))$ is a bounded poset where the smallest graph $(V,\emptyset)$ is the graph without edges and the greatest graph $(V,V\times V)$ is the graph in which any pair of nodes is adjacent. Therefore, the theory developed in Section 6.1 can be used to aggregate random graphs.

One of the most known models for random graphs is the so-called Erdős-Rényi model [123]. The random graph is defined on a fixed V, and the edges belong independently to E with probability p. A random Erdős–Rényi graph will be denoted as $(V,E(p))$. Considering the usual stochastic order defined in Definition 6.3, it is immediate that $(V,E_1(p_1)) \leq_{st} (V,E_2(p_2))$ if and only if $p_1 \leq p_2$. For a general random graph, it will be denoted as (V,E_ω), being E_ω a function with Ω as its domain. In the following example two of the most natural (monotonic) aggregation of random graphs are given, the union and intersection.

Example 6.60 *[31] Let $U : L^n_{\mathscr{G}} \to L_{\mathscr{G}}$ be the function such that, given the random graphs $(V,E_{1,\omega}),\ldots,(V,E_{n,\omega}) \in L^n_{\mathscr{G}}$,*

$$U\left((V,E_{1,\omega}),\ldots,(V,E_{n,\omega})\right) = \left(V,\cup_{i=1}^n E_{i,\omega}\right).$$

Also consider the function $I : L^n_{\mathscr{G}} \to L_{\mathscr{G}}$ such that:

$$I\left((V,E_{1,\omega}),\ldots,(V,E_{n,\omega})\right) = \left(V,\cap_{i=1}^n E_{i,\omega}\right).$$

Both are aggregations of random graphs, since they can be induced by the union and intersection of deterministic graphs (see [122]). In particular, if n independent Erdős–Rényi graphs are considered, then the distribution of the aggregation can be easily proven to be:

$$U\left((V,E_1(p_1)),\ldots,(V,E_n(p_n))\right) =_{st} \left(V,E\left(1-\prod_{i=1}^n (1-p_i)\right)\right),$$

$$I\left((V,E_1(p_1)),\ldots,(V,E_n(p_n))\right) =_{st} \left(V,E\left(\prod_{i=1}^n p_i\right)\right).$$

6.5.3 Aggregation of random positive semi-definite matrices

Random positive semi-definite matrices appear mainly in the estimation of covariance matrices (see [278]). One of the most common distributions for random matrices is the Wishart distribution [323], associated with the estimation of the covariance matrix of a multivariate Gaussian random vector. Aggregation of random matrices can appear, for instance, in the covariance estimation when considering Multivariate Analysis of Variance (MANOVA) [70], in which, under certain conditions, the covariance estimates in the different groups are fused into a unique estimation.

The usual order defined over positive semi-definite matrices is the Loewner order, which evaluates the positive semi-definiteness of the difference of the matrices. Denote the set of all positive semi-definite (of fixed dimension n) as S_n^+.

Definition 6.61 *[314] Let $A, B \in S_n^+$ be two positive semi-definite matrices. Then, A is smaller than or equal to B with respect to the Loewner order if $B - A \in S_n^+$.*

The latter relation is a partial order. It should be noted that the infimum of S_n^+ is the null matrix, but S_n^+ does not have a supremum. The definition of a stochastic order for random positive semi-definite matrices can be done with techniques similar to the ones introduced in Section 6.1. In the next example, some functions that are increasing with respect to the Loewner order are provided.

Example 6.62 *[31] Let M_1 and M_2 be two positive semi-definite matrices. Then, the following operations are increasing with respect to the Loewner order.*

- $A(M_1, M_2) = 0.5M_1 + 0.5M_2$,
- $A(M_1, M_2) = tr(M_1)M_2$,
- $A(M_1, M_2) = |M_1|M_2$,
- $A(M_1, M_2) = \begin{vmatrix} (M_1)_{\{1,\ldots,k\}} & M_0 \\ M_0^t & (M_2)_{\{k+1,\ldots,n\}} \end{vmatrix}$, *where $k \leq n$, $(M_1)_{\{1,\ldots,k\}}$ is the first diagonal block of dimension $k \times k$ of M_1, $(M_2)_{\{k+1,\ldots,n\}}$ the last diagonal block of dimension $n-k \times n-k$ of M_2, M_0 is a $k \times n-k$ null matrix and M_0^t is its transpose.*

Notice that, with the appropriate construction of the stochastic order, the latter functions applied to random matrices could be monotonic in the stochastic sense. However, in order to work with aggregations of random positive semi-definite matrices, the upper boundary condition should be studied in detail. An easy solution could be to impose a diagonal matrix with large diagonal elements as an upper bound. Another possibility is to use arguments similar to those in Theorem 6.15 to deal with the unboundedness of S_n^+. In any case, such a study will not be disclosed here.

6.5.4 Aggregation of fuzzy random variables

Fuzzy random variables are random elements from a probability space to the space of fuzzy numbers of $\mathbb{R}^n$. More precisely, the set $F_0(\mathbb{R}^n)$ is defined as the set of all functions $u : \mathbb{R}^n \to [0,1]$ such that $\{\vec{x} \in \mathbb{R}^n \mid u(x) \geq \alpha\}$ is non-empty and compact for each $\alpha \in (0,1]$, which is associated with all fuzzy numbers on $\mathbb{R}^n$ [272]. The elements of $F_0(\mathbb{R}^n)$ are functions that give a membership degree between 0 and 1 of the fuzzy number to each of the elements of $\mathbb{R}^n$. Recall that $\mathbb{B}^n$ denotes the Borel σ-algebra associated with $\mathbb{R}^n$. The formal definition of fuzzy random variable is the following one.

Definition 6.63 *[272] Let $(\Omega, \mathscr{F}, P)$ be a probability space. A fuzzy random variable is a function $X : \Omega \to F_0(\mathbb{R}^n)$ such that*

$$\{(\omega, \vec{x}) \in \Omega \times \mathbb{R}^n \mid \vec{x} \in X_\alpha(\omega)\} \in \mathscr{F} \times \mathbb{B}^n,$$

for every $\alpha \in [0,1]$, where $X_\alpha : \Omega \to \mathscr{P}(\mathbb{R}^n)$ is defined by

$$X_\alpha(\omega) = \{\vec{x} \in \mathbb{R}^n \mid X(\omega)(\vec{x}) \geq \alpha\}.$$

Fuzzy random variables are used for combining two sources of uncertainty, randomness and imprecision, the first being associated with the difference between two measurements of the quantity and the second to a vagueness of its value. For some examples of their application, the reader is referred to [94, 126]

Fuzzy random variables can be seen with three different perspectives. Firstly, they can be seen as random elements over the space $F_0(\mathbb{R}^n)$. Notice that several

orders can be defined between the elements of $F_0(\mathbb{R}^n)$ (see [339]). In that direction, if instead of $\mathbb{R}^n$ the Cartesian product of a bounded interval I is considered, one could think $F_0(I^n)$ as a bounded poset. Therefore, the approach of Section 6.1 can be used to define aggregations of fuzzy random variables in this regard.

Secondly, notice that, as stated in [94], the function X_α is a random set for any $\alpha \in [0,1]$ when considering the Borel σ-algebra associated with the topology generated by the Hausdorff metric [159]. This notion, see Theorem 2.7 in [239], is equivalent to Effros measurability. Then, one could define a usual stochastic order for fuzzy random variables as follows.

Definition 6.64 *Let X and Y be two fuzzy random variables. Then, X is said to be smaller than or equal to Y in the usual stochastic order for fuzzy random variables if $X_\alpha \leq_{st} Y_\alpha$ for any $\alpha \in [0,1]$.*

Of course, some properties of the introduced stochastic order should be studied. Moreover, to obtain a Composition Theorem for fuzzy random variables similar to Theorem 3.12 or Theorem 6.58, it will be necessary to deal with the Zadeh extension principle [107].

Finally, fuzzy random variables, since their values are functions in $F_0(\mathbb{R}^n)$, can be seen as stochastic processes indexed by $\mathbb{R}^n$ and having random variables taking values in the unit interval. However, big values of the function in $F_0(\mathbb{R}^n)$ do not imply that the fuzzy random variable is big, since its meaning has to be with membership degrees, not with its location. Therefore, the approach given in Section 6.3, although applicable, is not adequate in this case.

Chapter 7

Numerical results, simulations and real problems

Contents

The latter chapters have been devoted to developing the theory of aggregation of random structures in different ways. Focusing on a more experimental side of the topic, in this chapter the applicability of the introduced concepts is illustrated. In particular, some of the defined aggregation operators and introduced methods are being tested by considering both real databases and simulated studies.

Moving from the mathematical theory to its application to data is not straightforward, since sometimes the mathematical concepts are not easy to be represented properly in computers, where the computations have to been done. In this regard, some comments about the computability of the aggregation of random structures

are given in Section 7.1.

The experiments are related to some of the aggregations of random variables and associated results given in Chapter 4, as well as to an application of the stochastic comparison of variability measures given in Sections 5.2 and 5.3. In particular, using the results of Section 4.2, a flexible-dimensional estimator, i.e. with a non-fixed number of inputs, is first defined in Section 7.2. The estimator regards the mean of symmetric random variables. Subsequently, a simulation study considering eight relevant symmetric distribution has been done, illustrating the benefits of the method.

In Section 4.3, many properties of uniform capacities were proved. Some of them can be tested using usual hypothesis tests in the literature, so the uniformity of two of the main random generators of capacities is tested in this regard in Section 7.3.

Then, the SOWA operator, introduced in Section 4.4.3, is applied to the prediction of the temperature of a room considering the temperature of the rest of the rooms of the house in Section 7.4. Its behavior is compared with the weighted arithmetic mean and the OWA operator, with the proposal being the best of the three.

Section 7.5 is devoted to illustrate the importance of the results given in Sections 5.2 and 5.3, in addition to the study of the properties of the minimum of penalty functions. More precisely, the proved results are used to propose statistical tests for the multivariate convex order, the supermodular order and for the dispersive order (the latter one under some assumptions). The introduced methods are illustrated by small examples that show its good behavior.

Finally, the IOLF operator and related functions are used in a problem considering 18 time series of real data of temperature and humidity. As a first step, seven prediction models with different nature and characteristics are trained, and then six pre-aggregation functions are used as ensembles. As a result, it is stated that a modification of the weighted arithmetic mean, but with negative weights, is significantly the best alternative. These results can be found in Section 7.6.

7.1 Some comments about computability

In addition to the clear theoretical interest, the aggregation of random structures is also devoted to its application to data analysis. In this respect, when dealing with real observations, it is necessary to use numerical methods and computers to obtain the results.

However, all the constructions given in the previous sections are theoretical mathematical concepts that are not easily represented in a computer. Focusing on the simplest example, random variables, even their representation in a computer environment, is not straightforward. A detailed study of such a representation can be found in [28], while only a summary is given here.

The first alternative to represent a random variable on a computer is a function that implements the mathematically measurable function from a probability space to real numbers. In particular, a fixed probability space $(\Omega, \mathscr{F}, P)$ has to be stored on the computer and then a computer function (X) that maps any value of Ω to $\mathbb{R}$ and satisfies the measurability property has to be considered. This is the most faithful representation of a random variable, the implementation of the function itself, but it entails several problems. The first is that the measurability condition is not easy to check in a computer program. This can be solved by considering $\mathscr{F}$ as the parts of Ω, $\mathscr{P}(\Omega)$, so any function is measurable. In addition, it is not clear how to model a usual probability space as the unit interval with the Borel σ-algebra and the Lebesgue measure. Even if one could solve the latter problems, the space L_I^n has an infinite number of elements.

The second alternative is to represent the random variable by means of implementing its cumulative distribution. This makes sense for a wide collection of applications in which only the distribution of the random variable is relevant, but not the underlying structure of a measurable function from a probability space to the real numbers. In addition, it is quite simpler than in the previous case. Notice also that, as stated in Lemma 3.7, the distribution of the output of the aggregation of random variables can be determined by the distribution of the input. The main drawback of this representation is that there exist different random variables with the same distribution function. Another problem is that the possible dependence between some random variables cannot be considered just by specifying their marginal

distribution functions. Moreover, the computation of the distribution of the output is not straightforward for continuous random variables, since numerical integration is needed in most cases.

The last alternative is to identify random variables with generators of pseudo-random numbers. This case is the farthest from the initial concept of a random variable, since the probability structure of the random variables is unknown, only observations of them are available. However, in some cases, this approach suffices for applied purposes. In particular, a simulation can be a good approximation of the behavior of a good number of random systems [124, 179]. Furthermore, the distribution of a combination of random variables is generally not easy to compute explicitly, so the use of simulation eases the computations. Of course, a distribution function can always be approximated by empirical distribution functions [297].

Leaving aside the representation of the random variables, the reality is that, in many cases, it is not necessary to know the image of each of the elements in L_I^n, not even its distribution. For problems regarding prediction or estimation, the given data are always observations of the random elements, that is, the value of the random elements for a particular and unknown element of the probability space. For instance, given an aggregation of random variables A, it is often enough to know the distribution of $[A(\vec{X}) \mid \vec{X} = \vec{x}]$.

In this direction, induced aggregations are very convenient, since $[A(\vec{X}) \mid \vec{X} = \vec{x}] = \hat{A}(\vec{x})$ where $\hat{A}$ is the aggregation that induces A. Another simple case in this regard is the case of conditionally determined aggregations, in which $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ has always degenerate distribution, thus the result is just a real number. Notice that, even if the determination of this real number depends on the random vector $\vec{X}$, the typical case is to have a dependence with respect to a quantity associated with the distribution of $\vec{X}$ such as its mean vector. These parameters can be estimated with the available data, see Section 7.4 for an applied example in this regard, so this is not a main drawback for its applicability.

If the aggregation is not conditionally determined, there are still some cases in which $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ can be easily determined. For randomly induced aggregations, the case in which the family of random parameters $(\lambda_{\vec{X}}, \vec{X} \in L_I^n)$ fulfills that $\vec{X}$ and $\vec{\lambda}_{\vec{X}}$ are independent for any $\vec{X} \in L_I^n$ is especially simple. Computing the distribution of $[A(\vec{X}) \mid \vec{X} = \vec{x}]$ to, for instance, determine confidence intervals, is easy when

knowing $\vec{x}$ by using simulation.

Finally, some particular cases of aggregations of random variables with the same distribution as an induced one can be easily computed. Given two aggregations of random variables $A, B : L_I^n \to L_I$ with the same distribution, if the marginal distribution functions and the copula of $(A(\vec{X}), B(\vec{X}))$ are known, then the distribution of $[B(\vec{X}) \mid A(\vec{X}) = x]$ can be simulated using the techniques given in Chapter 2 in [253].

7.2 Flexible-dimensional L-statistics for mean estimation of symmetric random variables

The study of the population mean of a random variable is a central problem in Statistics [278]. Several methods for its estimation have been developed, some of them focused on order statistics [218]. In particular, the use of linear combinations of them, known as L-statistics, has been deeply studied in the specialized literature [2, 121, 138, 157, 203].

In some approaches, the weights of such a linear combination are computed by integrating a weight-generating function $h : [0,1] \to \mathbb{R}$ [164]. If the dimension of the weighting vector is n, the $k-th$ weight is computed as:

$$w_k = \int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t)dt,$$

for each $k \in [n]$.

A similar approach in Aggregation Theory, usually called Yager's method to derive Ordered Weighted Averaging (OWA) weights [140, 327], consists of considering an increasing bijection $g : [0,1] \to [0,1]$ and computing these weights as

$$w_k = g\left(\frac{k}{n}\right) - g\left(\frac{k-1}{n}\right),$$

for each $k \in [n]$.

From the mean estimation point of view, the weights that minimize the Mean Squared Error are especially interesting. Although these MSE-minimizing weights

may be computed if the distribution belongs to a scale-location family, i.e. random variables fulfilling $X = \lambda Y + \mu$ for a fixed random variable Y and $\mu, \lambda \in \mathbb{R}$, or if there are some available data, the resulting weights are only defined for a particular sample size. Nevertheless, in real-world problems, the sample size may change and therefore the optimal weights are no longer applicable. The most prominent example is the case of censored samples [6, 8, 250], which appear naturally in survival analysis [189], missing observations [214] or the changing sampling frequency that may appear in signal analysis [17]. It is therefore essential to provide a method that allows the sample size to be modified in order to deal with real-world scenarios.

In this sense, given a distribution, it would be convenient to be able to find a certain generating function g, as the one used in Yager's method [327], to derive a weighting vector of the required size. However, given a distribution, it is not easy, in general, to find such a function g to generate the optimal weights, in the sense of minimizing the MSE in mean estimation. Only some distributions, such as the Gaussian and the uniform distributions, have a simple pattern for the optimal weights that allows their computation from a generating function [72].

Theorem 4.12, the numerical results shown in Figure 4.1 and the performance of the EVR-OWA operator [139, 140] to provide symmetrically ordered aggregations, serve as inspiration for a method to construct a flexible-dimensional L-statistic for mean estimation.

In particular, given some optimal weights, the associated cumulative weights are fitted using a function $g : [0,1] \to \mathbb{R}$. Then, if a vector with different dimension is aggregated, the fitted function is used to generate new weights that suit the new dimension. Keeping in mind Theorem 4.12 and Figure 4.1, it is reasonable to think that the generated weights will be similar to the optimal weights if both dimensions are sufficiently big and close.

This method can be used in two different scenarios. In the first one, the expression of the underlying distribution is known, but it is complicated to derive the optimal weights analytically. This is a common situation, since the distributions of L-statistics are usually hard to handle (see Chapter 6 in [104]). By simulation, it is possible to compute a good approximation of the optimal weights. However, it is necessary to perform such a simulation for each sample size and, for big sample sizes, the computational time can be unfeasible. In this regard, the method proposed

in this section allows one to obtain an approximation to derive the optimal weights in a simple way for different sample sizes, especially for big ones.

In the second one, a quantity of interest is supposed to be independently measured when it is perturbed by an additive symmetric noise with a mean of 0. For each of the values of this quantity, several measures are made. Given a dataset in which there are different true values of the quantity and their associated measurements, it is possible to fit an L-statistic that minimizes the MSE in the dataset. Then, the constructed L-statistic can be applied to new data to obtain estimations of the quantity of interest. However, if the sample size changes, the fitted weights are no longer valid, since they were computed for a fixed length. In this sense, for small variations of sample sizes, the method proposed in this section allows obtaining new weights that will lead to an estimator with a small MSE and defined with the correct dimension.

Although the assumption of Theorem 4.12 may be difficult to check for an arbitrary distribution, the presented method is illustrated for several classical symmetric distributions and shows a good behavior.

7.2.1 Fitting functions

Even if the limit function is known, the optimal cumulative weights can be far from it for small sample sizes. Therefore, one of the main challenges faced by this procedure is the correct choice of a family of functions to fit the cumulative weights. Although another choice might be done, it will be considered a family of functions based on Extreme Value Amplifications (EVAs) [139] and Extreme Value Reductions (EVRs) [140] due to their convenient behavior when applied in OWA aggregations [36]. In particular, the following families of functions are considered:

- Sin-based EVAs/EVRs: $s_\alpha : [0,1] \to [0,1]$ defined as

$$s_\alpha(x) = x + \sum_{k=0}^{n} \alpha_k \sin(2\pi k x) \ \forall x \in [0,1],$$

  with $(\alpha_k, k \in \mathbb{N} \cup \{0\})$ being a sequence,

- Grade 3 polynomials: $p^3_\beta : [0,1] \to [0,1]$, defined as

$$p^3_\beta(x) = (1-\beta)x + 3\beta x^2 - 2\beta x^3 \ \forall x \in [0,1],$$

  with $\beta \in [-1,1]$,

- Spline-based EVAs and EVRs: $sp_\gamma : [0,1] \to [0,1]$ defined as

$$sp_\gamma(x) = \begin{cases} \frac{1}{2} - \frac{1}{2}(1-2x)^\gamma & \text{if } 0 \leq x < \frac{1}{2}, \\ \frac{1}{2} + \frac{1}{2}(2x-1)^\gamma & \text{if } \frac{1}{2} \leq x \leq 1, \end{cases}$$

  with $\gamma \in (0,\infty)$,

- Pseudo-constant function: $c : [0,1] \to [0,1]$ defined as

$$c(x) = \begin{cases} 0 & \text{if } x = 0, \\ \frac{1}{2} & \text{if } 0 < x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

For the family s_α, which depends on an infinite number of parameters, it will just be considered a vector of dimension 4, $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

Then, a linear combination of the aforementioned families with coefficients $\vec{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is considered to obtain a unique function with better behavior. Therefore, define the family of functions $g_{\vec{\alpha},\beta,n,\vec{\lambda}} : [0,1] \to \mathbb{R}$ given by

$$g_{\vec{\alpha},\beta,\gamma,\vec{\lambda}}(x) = \lambda_1 s_{\vec{\alpha}}(x) + \lambda_2 p^3_\beta(x) + \lambda_3 sp_\gamma(x) + \lambda_4 c(x).$$

7.2.2 Numerical results

For eight different symmetric distributions, the optimal weights for $n = 20$ are computed through simulation. Then, the cumulative weights are fitted using the latter family of functions, with the aim to minimize the Mean Squared Error between the points and the fitted function. The resulting function is compared to the optimal weights for $n \in \{18, 19, 21, 22\}$, the dimensional sizes closest to 20 and $n \in \{10, 30\}$, further ones. The choice of $n = 20$ is adequate to illustrate the method working on small sample sizes, in which small differences between sample sizes

are relevant. Additional experiments with larger sample sizes have also been conducted, with better results as a consequence of the convergence of Theorem 4.12. For too small samples, the differences between sample sizes are relatively large, and the behavior gets worse.

In particular, the considered distributions are the Laplace distribution, the hyperbolic secant distribution, the Student's t-distribution (with 30 degrees of freedom), two Generalized Normal (or G-normal) distributions (with parameters $s = 3$ and $s = 1.5$), two Beta distributions (with $\alpha = \beta = 0.5$ and $\alpha = \beta = 2$) and the logistic distribution. They are a good sample of classical and relevant distributions in theoretical and applied problems in Statistics. The uniform and the normal distribution have not been addressed because the optimal weights are straightforward to compute for any sample size, $\left(\frac{1}{2}, 0, \ldots, 0, \frac{1}{2}\right)$ and $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ [278].

It is important to remark that the here-proposed method is not restricted to the distributions considered in this section. As guaranteed by Theorem 4.12, for a sufficient regular distribution, this procedure can also be used, but perhaps using a wider family of fitting functions.

The resulting parameters of the fitted functions can be found in Table 7.1. If a coefficient of the linear combination is 0, the parameters of the associated function are not provided.

Distribution	λ_1	λ_2	λ_3	λ_4	α_1	α_2	α_3	α_4	β	γ
Laplace	1	0	0	0	-0.272	0.084	-0.026	0.008	-	-
Hyperbolic secant	1	0	0	0	-0.157	0	0	0	-	-
Student's t	0.923	0	0.087	-0.010	-0.069	-0.015	-0.005	-0.004	-	0.870
G-normal $s = 3$	0.313	0.176	0.485	0.026	0.137	0.017	0.008	0.004	0	2.697
G-normal $s = 1.5$	1	0	0	0	-0.271	0.084	-0.026	0.008	-	-
Beta $\alpha = \beta = 0.5$	0	-0.070	-0.320	1.390	-	-	-	-	-1.857	13.04
Beta $\alpha = \beta = 2$	0.046	0.352	0.159	0.443	0.208	0.087	0.051	0.034	-0.492	5.738
Logistic	0.584	0.416	0	0	-0.099	0.012	-0.004	-0.001	1.022	-

Table 7.1: Parameters of the fitted function to the optimal cumulative weights of the considered distributions with a sample size $n = 20$ [37].

The limit function for the logistic distribution, derived in Example 4.14 equals $3x^2 - 2x^3$, and the fitted function, using the optimal weights for $n = 20$, is $0.584(x - 0.099\sin(2\pi x - \pi) + 0.012\sin(4\pi x - \pi) - 0.004\sin(6\pi x - \pi) - 0.001\sin(8\pi x -$

$\pi)) + 0.416(0.022x + 3.066x^2 - 2.044x^3)$. Although the last part of the fitted function is similar to the limit function, there is a notable difference between them. For bigger sample sizes, the fitted function will converge to the limit function, as already proved in Theorem 4.12. For the hyperbolic secant distribution, the fitted function is almost equal to the one computed in Example 4.13.

In Figure 7.1, the fitted functions and the cumulative weights can be seen for different values of n, for all the considered distributions. It can be seen that the fitted function also serves as a good approximation for the cumulative weights when $n \in \{18, 19, 21, 22\}$, although they have not been used to fit the function. As expected, the behavior is worse for the cases $n \in \{10, 30\}$.

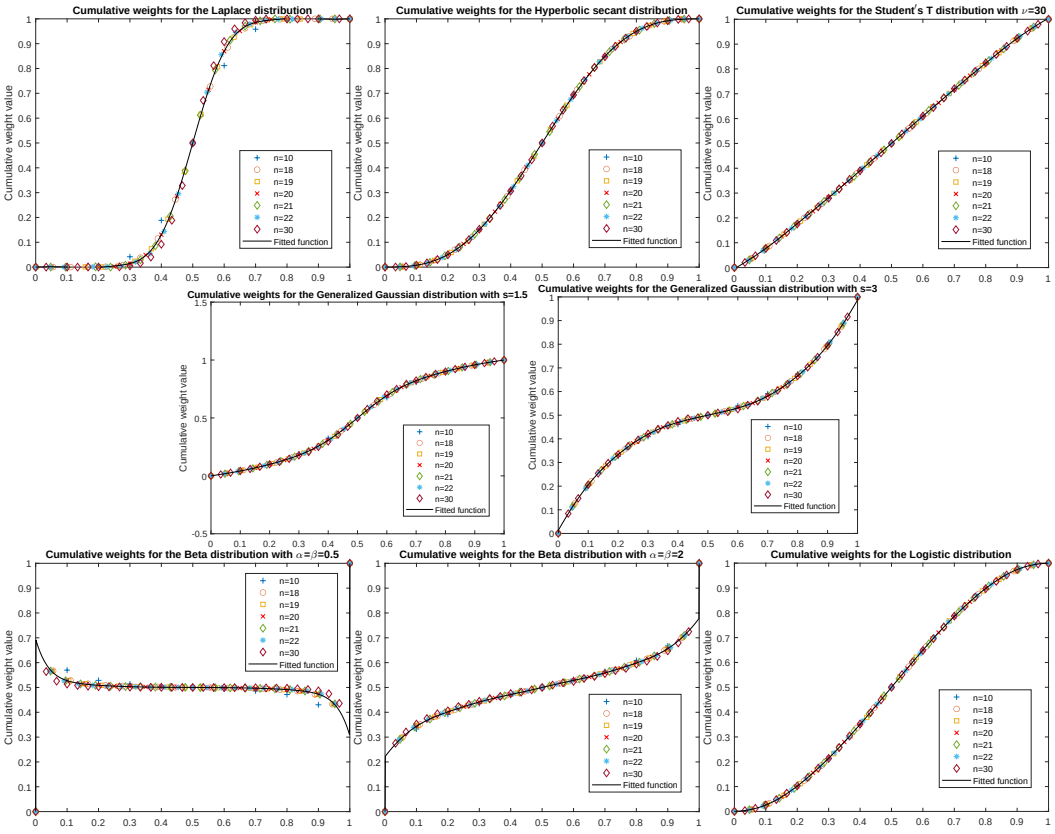

Figure 7.1: Fitted cumulative weights when $n = 20$ and their comparison with the cumulative weights when $n \in \{10, 18, 19, 21, 22, 30\}$ for some symmetric distributions [37].

In addition, the Mean Squared Error between the points and the fitted function have been computed for all the cases. In Table 7.2, it can be seen that the MSE does not increase considerably when moving from $n = 20$ to closer sample sizes (even in a particular case it decreases) and it is of the order of $10^{-6} - 10^{-8}$ (notice that the values in the table are multiplied by 10^{-7}) for all cases. The MSE for the other sample sizes increases in some cases to the order of $10^{-4} - 10^{-6}$.

Distribution	$n = 10$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$	$n = 30$
Laplace	746.8	17.52	6.830	2.384	12.90	25.58	309.7
Hyperbolic secant	13.47	2.870	1.280	0.304	1.656	0.988	19.30
Student's t	9.074	2.374	4.071	1.245	1.723	2.260	14.28
G-normal $s = 3$	435.2	0.680	1.886	0.177	1.420	1.313	42.06
G-normal $s = 1.5$	992.0	40.34	21.38	3.322	9.794	32.83	249.0
Beta $\alpha = \beta = 0.5$	4267	0.540	0.639	0.356	0.696	2.395	1131
Beta $\alpha = \beta = 2$	551.1	9.106	1.576	0.530	1.806	6.971	248.6
Logistic	11.25	11.80	16.80	2.200	7.903	7.273	8.947

Table 7.2: Mean Squared Error (multiplied by 10^{-7}) between the fitted function using the cumulative weights for $n = 20$ and the cumulative weights for $n \in \{10, 18, 19, 20, 21, 22, 30\}$ for the considered distributions [37].

The small difference between optimal and fitted weights should lead to a small difference in the behavior between the obtained L-estimators. In this regard, the MSE with respect to the real value of the mean, also considering the arithmetic mean, has been computed by simulation. The results can be found in Table 7.3.

As can be seen, the L-estimators with fitted weights behave similarly to the optimal one. Their MSE is always between the optimal, which cannot be improved, and the sample mean, which can be seen as a naive flexible-dimensional method. In fact, it is almost the same for most of the considered cases. The difference is greater for sample sizes that are far from $n = 20$, since, as already seen in Table 7.2, the fitted function is a better approximation for closer sample sizes.

In order to reproduce the latter results for non-symmetric distributions, it is necessary to prove Theorem 4.12 for non-symmetric distributions, which is far from being immediate. In addition, the family of functions $g_{\vec{\alpha},\beta,n,\vec{\lambda}}$ should be extended to

deal with non-symmetric weights. If both points can be solved, it will increase the applicability of the method, since, in many cases, real data do not have a symmetric distribution.

Distribution	Weights	$n=10$	$n=18$	$n=19$	$n=20$	$n=21$	$n=22$	$n=30$
Laplace	Optimal	0.1417	0.0720	0.0682	0.0652	0.0603	0.0558	0.0400
	Fitted	0.1422	0.0720	0.0682	0.0652	0.0603	0.0558	0.0401
	Balanced	0.1986	0.1091	0.1067	0.1011	0.0946	0.089	0.0656
Hyperbolic secant	Optimal	0.0869	0.0463	0.0445	0.0410	0.0408	0.0377	0.0281
	Fitted	0.0869	0.0463	0.0445	0.0410	0.0408	0.0377	0.0281
	Balanced	0.1010	0.0553	0.0530	0.0493	0.0485	0.0450	0.0343
Students T	Optimal	0.1070	0.0599	0.0566	0.0527	0.0503	0.0485	0.0344
	Fitted	0.1070	0.0599	0.0566	0.0527	0.0503	0.0485	0.0344
	Balanced	0.1073	0.0603	0.0567	0.0528	0.0506	0.0488	0.0345
G-normal $s=3$ ($\times 10^{-1}$)	Optimal	0.3454	0.1905	0.1831	0.1685	0.1584	0.1506	0.1139
	Fitted	0.3455	0.1905	0.1831	0.1685	0.1584	0.1506	0.1139
	Balanced	0.3730	0.2094	0.1995	0.1867	0.1724	0.1669	0.1268
G-normal $s=1.5$	Optimal	5.4916	2.7737	2.6216	2.4403	2.3005	2.2196	1.5294
	Fitted	5.7613	2.8039	2.6418	2.4585	2.3036	2.2265	1.5665
	Balanced	12.048	6.7415	6.4255	6.1515	5.8014	5.5608	4.0393
Beta $\alpha=\beta=0.5$ ($\times 10^{-2}$)	Optimal	0.1926	0.0294	0.0243	0.0193	0.0182	0.0137	0.0049
	Fitted	0.1982	0.0294	0.0244	0.0193	0.0182	0.0137	0.0052
	Balanced	1.2537	0.7083	0.6566	0.6329	0.5902	0.5664	0.4134
Beta $\alpha=\beta=0.5$ ($\times 10^{-2}$)	Optimal	0.3857	0.1939	0.1844	0.1748	0.1683	0.1559	0.1077
	Fitted	0.3859	0.1939	0.1844	0.1748	0.1683	0.1559	0.1078
	Balanced	0.4896	0.2742	0.2584	0.2509	0.2437	0.2291	0.1681
Logistic	Optimal	0.3116	0.1710	0.1594	0.1507	0.1431	0.1405	0.0990
	Fitted	0.3117	0.1710	0.1594	0.1507	0.1431	0.1406	0.0990
	Balanced	0.3341	0.1836	0.1727	0.1630	0.1542	0.1523	0.1076

Table 7.3: Mean Squared Error of L-estimators when considering the optimal weights, the fitted weights and the balanced weights for sample sizes $n \in \{10, 18, 19, 20, 21, 22, 30\}$ for the considered distributions [37].

7.3 Statistical tests for uniformity in generation of capacities

Many algorithms have been developed for the random generation of capacities or fuzzy measures [43, 44, 46, 51, 95, 96, 148, 167]. As briefly explained in Section 4.3, the main objective of such algorithms is to generate capacities uniformly over the set of all capacities (of a particular dimension) as fast as possible. The structure of the set of all capacities is difficult to deal with, since it is an order polytope of order $2^n - 2$ in which even the number of possible orderings of the values is hard to determine [96].

However, although uniformity is the main objective of these algorithms, there do not exist tools specifically designed for checking such uniformity in the literature. Some general algorithms for checking the uniformity over (multivariate) sets can be found in [56, 162, 169, 270, 301, 305].

They are typically constructed by computing an associate quantity to the sample points, for instance, the Minimum Spanning Tree, the length of the smallest Hamiltonian path, the mean distance between one point and its closest neighbor or the distance to the boundary. Then, the value of the quantity is compared with the one expected for the uniform distribution.

These algorithms have mainly two drawbacks. The first is that, in general, the distribution of quantity is not known. Therefore, it has to be simulated by considering a uniform generator. Then, for the approach considered in this section, it would be necessary to have a uniform generator of capacities in which uniformity is ensured to check the uniformity of a second algorithm. The second one is that, for a big dimension, such algorithms are very slow. Notice that the number of values of a capacity grows as 2^n with the dimension, thus for big dimensions such algorithms will take too much time.

In contrast, testing some of the properties already proved in Section 4.3 can be done without a prior uniform algorithm and can be computed in a reasonable amount of time. Which of them can be tested will be determined in the next section.

7.3.1 Testable properties

This section is devoted to discuss which properties for uniform random capacities among the ones proved in Section 4.3 can be tested, which statistical tests can be used and why the remaining properties are not possible to be tested. In the following, consider samples consisting of m vectors which are independent observations of a uniform random capacity.

The properties provided in Proposition 4.21 are easy to test. In particular, the equality of distribution for subsets of the same cardinality and the auto-duality in distribution of the uniform random capacity can be tested by applying 2-sample Kolmogorov-Smirnov tests. It is important to remark that the third point in the result can be derived from the latter tests, since if $|A| \leq |B|$, there always exist A', B' such that $|A| = |A'|$, $|B| = |B'|$ and $A' \subseteq B'$, thus $\mu(A) =_{st} \mu(A') \leq \mu(B') =_{st} \mu(B)$. The first property and Corollary 4.22 can be tested using an exchangeability test, but they are very expensive in terms of computation time.

The measure of subsets following a mixture of unknown Beta distributions cannot be tested, since such a test does not exist in the literature. For the symmetry of the Orness, it is possible to use the Cabilio-Masaro symmetry test and for its mean (and median) being 0.5, a Wilcoxon test can be performed.

In addition, if the values of each of the m vectors of the sample are merged into one vector, eliminating in each of them the measure of the empty set and the whole set (which are always 0 and 1), $m(2^n - 2)$ values that consist of a particular ordering of a simple random sample of a uniform distribution over $[0, 1]$ are obtained. Then, a 1-sample Kolmogorov-Sminorv test can be applied to test the uniformity. Notice that it is not necessary to reorder the sample before applying the test, since in the Kolmogorov-Smirnov test the sample is ordered from the smallest to the greatest value to construct the empirical distribution function.

Other properties may be tested when restricting the study to balanced, belief or possibility measures. In that case, it is necessary to consider the observations that belong to each of the families and then, using the associated subsamples, test the properties. Numerical results show that the probability of a uniform random capacity to be a belief or possibility measure is 0, thus the related subsamples are empty and no statistical test can be used.

For balanced measures, around 25% for $n = 4$ and 1.3% for $n = 5$ of the observations of a uniform random capacity are balanced measures, thus their properties can be tested. In particular, the theoretical distribution of the measure of the subsets, the Orness and the non-modularity index of the whole set are known. For the first two cases, the distribution functions can be constructed and then 1-sample Kolmogorov-Smirnov tests can be performed. In the last one, the expression is quite complicated and includes a derivative that, even for small values of n, is hard to compute.

7.3.2 Uniformity study of random capacity generation algorithms

In this section, the statistical tests explained in the last section are applied to two algorithms of random generation of capacities, to determine whenever the generated values follow a uniform distribution or not.

The two algorithm are included among the most used in the literature. They are called randomized topological sort (tsort) [51] and minimum-plus (minplus) [95], which are implemented in the R package Rmftool version 4.1.10 [42] in the functions $fm.generate\_fm\_tsort$ and $fm.generate\_fm\_minplus$, respectively. Both functions also have an optional parameter *rejection*, that forces the algorithm to reject linear extensions that have been seen already. Using this option, the algorithm tends not to repeat choosing the same linear extension more than a few times. The cases of dimension $n = 4$ and $n = 5$, with sample sizes $k = 1000$ and $k = 10000$, with and without the rejection option, are considered for the study.

The tested null hypothesis are the following:

- μX: The measures of subsets of cardinality X have the same distribution,
- DX: The distribution of the measure and the dual measure of the subsets of cardinality X are the same,
- $MOrn$: The mean of the Orness is 0.5,
- $SOrn$: The distribution of the Orness is symmetric,
- $Dist$: The distribution of the distance to the closest 0-1 capacity is the one given in Proposition 4.27,

Algorithm	rejection	sample size	$\mu 1$	$\mu 2$	$\mu 3$	$D1$	$D2$	$MOrn$	$SOrn$	$Dist$	$Unif$	$B\mu 1$	$B\mu 2$	$B\mu 3$	$BOrn$
tsort	N	1000	0.655	0.613	0.643	0.497	0.491	0.443	0.508	0.450	0.536	0.546	0.520	0.510	0.485
	Y	1000	0.643	0.608	0.665	0.503	0.462	0.497	0.505	0.507	0.539	0.524	0.521	0.495	0.475
	N	10000	0.652	0.618	0.657	0.513	0.486	0.519	0.502	0.512	0.489	0.515	0.509	0.499	0.287
	Y	10000	0.645	0.609	0.645	0.496	0.495	0.538	0.516	0.507	0.514	0.513	0.508	0.509	0.320
minplus	N	1000	0.663	0.622	0.622	0	0	0	0.487	0	0	0.329	0.065	0	0
	Y	1000	0.660	0.636	0.670	0	0	0	0.493	0	0	0.347	0.067	0	0
	N	10000	0.642	0.624	0.669	0	0	0	0.500	0	0	0.008	0	0	0
	Y	10000	0.642	0.624	0.669	0	0	0	0.500	0	0	0.008	0	0	0

Table 7.4: Average p-values among 100 repetitions for the considered algorithms of uniform generation of capacities of dimension $n = 4$ [27].

Algorithm	rejection	sample size	$\mu 1$	$\mu 2$	$\mu 3$	$\mu 4$	$D1$	$D2$	$MOrn$	$SOrn$	$Dist$	$Unif$	$B\mu 1$	$B\mu 2$	$B\mu 3$	$B\mu 4$	$BOrn$
tsort	N	1000	0.641	0.584	0.575	0.637	0.509	0.493	0.508	0.508	0.513	0.509	0.525	0.502	0.531	0.490	0.497
	Y	1000	0.627	0.588	0.596	0.620	0.510	0.495	0.506	0.507	0.512	0.514	0.497	0.507	0.498	0.502	0.479
	N	10000	0.612	0.582	0.583	0.614	0.501	0.486	0.500	0.488	0.501	0.523	0.506	0.515	0.526	0.528	0.421
	Y	10000	0.612	0.568	0.569	0.615	0.500	0.482	0.516	0.471	0.497	0.498	0.513	0.512	0.512	0.512	0.420
minplus	N	1000	0.652	0.579	0.594	0.637	0	0.004	0	0.458	0	0	0.450	0.484	0.484	0.190	0.373
	Y	1000	0.638	0.591	0.583	0.620	0	0.005	0	0.504	0	0	0.519	0.510	0.437	0.025	0.344
	N	10000	0.616	0.571	0.583	0.626	0	0	0	0.541	0	0	0.111	0.414	0.222	0	0.018
	Y	10000	0.612	0.579	0.584	0.633	0	0	0	0.464	0	0	0.107	0.420	0.213	0	0.038

Table 7.5: Average p-values among 100 repetitions for the considered algorithms of uniform generation of capacities of dimension $n = 5$ [27].

- $Unif$: The sample made by combining all the vectors of the sample and eliminating the measures for the empty set and the whole set is a rearrangement of a simple random sample of a uniform distribution,
- $B\mu X$: In the subsample consisting of balanced capacities, the distribution of the measure of subsets of cardinality X is the one given in Proposition 4.31,
- $BOrn$: In the subsample consisting of balanced capacities, the Orness has the distribution given in Proposition 4.32.

For each case, the procedure was repeated 100 times and the mean p-value was computed. The results for dimension 4 and 5 can be found in Tables 7.4 and 7.5.

As can be seen in both tables, the tsort algorithm seems to have a good behavior, since the average p-value in all the cases is above the usual significance levels (0.05, 0.01), therefore, the null hypothesis seems to not be rejected in any of the cases. Looking with more detail the p-values, it can be seen that they are smaller for the test of the distribution of the Orness when working with the subsample re-

lated consisting of balanced observations, but there is not enough evidence to have a rejection. Then, it is concluded that, with the tools developed in this thesis, there is no evidence to reject the uniformity of the samples produced by tsort, with and without the rejection option.

On the contrary, there is evidence of non-uniformity for minimal-plus algorithm. In particular, when the sample size is $k = 1000$ and $n = 4$, the hypothesis about the equality in distribution between the measure of subsets with the same cardinality and the symmetry of the Orness are rejected. Similar results are achieved when considering $n = 5$. Therefore, it is concluded that minimal-plus (or at least its implementation in Rfmtool 4.1.10) does not generate capacities with a uniform distribution, with and without the rejection option enabled.

7.4 Temperature prediction using the SOWA operator

This section is devoted to illustrate the advantages of the SOWA operator introduced in Section 4.4.3. Recall that this operator depends on a permutation π and a weighting vector $\vec{w}$. Firstly, the distribution functions $F_1, \ldots, F_n$ of the inputs are pointwise ordered to obtain the distribution functions $F_{[1]}, \ldots, F_{[n]}$ (see Figure 4.2), which are ordered with respect to the usual stochastic order. Then, the permutation π assigns the variability of each input to one of the constructed distribution functions and, finally, a weighted arithmetic mean of the resulting random variables is performed.

In the following, 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 $\hat{Y}$. Three different alternatives will be considered.

(1) $\hat{Y} = \sum_{i=1}^{n} w_i X_i$,

(2) $\hat{Y} = \sum_{i=1}^{n} w_i X_{(i)}$,

(3) $\hat{Y} = SOWA_{\vec{w},\pi}(\vec{X})$,

which are a linear combination of, respectively, the random variables, the order statistics and the components of the rearrangement increasing stochastically ordered

random vector introduced in Definition 4.42.

In particular, the database Appliances Energy Prediction, which can be found in [79], has been used. The considered data consist of the value of the temperature (in Celsius) in different locations of a house, measured in 19735 different times. The reader is referred to [80] for more details on the dataset. Dismissing the values associated with outside locations (since its temperature is quite different from the temperature inside the house), there are data of the temperature for the *kitchen, living room, laundry room, bathroom, teenager room, ironing room, parents room and office room*. The aim is to predict the temperature in the *office room* using the values in the rest of the indoor locations, considering the latter models. The variables are ordered in the vector in the order in which they have been enumerated above.

The dataset is divided into 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, it is firstly necessary to build the rearrangement increasing stochastically ordered random vector. Since the population distributions of the variables are unknown, the empirical distribution functions (see Definition 2.123) of the variables can be used as $F_1, \ldots, F_n$ in Definition 4.42. Then, the minimization is made on the set of all possible permutations π (which are $7! = 5040$) and the set of weighting vectors. The obtained optimal parameters for all the models are the following:

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

For the weighted arithmetic 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 close to 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 variability of the temperature in the *teenager room* is associated with the greatest distribution, and therefore, with a null weight. This coincides with the weight associated with the *teenager room* in model (1).

The Mean Squared Error, the Mean Absolute Error, and the Percentage Error, see [243], of the three models in the test sample are provided in Table 7.6. As can be seen, the SOWA has a better behavior than the other alternatives.

Model	MSE	MAE	PE (%)
Weighted Arithmetic Mean	0.28	0.44	1.93
Ordered Weighted Averaging	0.31	0.46	1.97
Stochastically Ordered Weighted Averaging	0.22	0.38	1.66

Table 7.6: Mean Squared Error (MSE), Mean Absolute Error (MAE) and Percent Error (PE) for the considered prediction models in the test sample [39].

These results are sensitive to the choice of the training and test samples. Therefore, the sample has been divided into 10 blocks of the same size and a cross-validation procedure (see [85]) 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. The average of the errors can be found in Table 7.7. It can be seen that the average errors are greater than in the first study. However, SOWA operator still has the better behavior among the alternatives.

Additionally, paired Wilcoxon tests have been applied to each pair of models, for each error measure. The results can be found in Table 7.8, in which it can be seen that the SOWA significantly overcomes the other alternatives in all the considered cases. In particular, since all the p-values associated with the alternative hypothesis *the SOWA has an error smaller than the other model* are below 0.05, the alternative hypothesis is accepted. Notice also that the OWA operator seems to be better than the WAM, but not always with enough significance.

Note that other aggregations, not only weighted averages, may be considered in

Model	MSE	MAE	PE (%)
Weighted Arithmetic Mean	0.86	0.73	3.61
Ordered Weighted Averaging	0.67	0.65	3.17
Stochastically Ordered Weighted Averaging	0.58	0.58	2.83

Table 7.7: Average Mean Squared Error (MSE), Mean Absolute Error (MAE) and Percent Error (PE) for the considered prediction models in the cross-validation procedure [39].

	Mean Squared Error			Mean Absolute Error			Percentage Error		
	WAM	OWA	SOWA	WAM	OWA	SOWA	WAM	OWA	SOWA
WAM	-	0.958	0.997	-	0.962	0.995	-	0.976	0.993
OWA	0.053	-	0.991	0.049	-	0.991	0.032	-	0.986
SOWA	0.005	0.012	-	0.007	0.012	-	0.009	0.019	-

Table 7.8: P-values for the Wilcoxon paired test with alternative hypothesis: the model of the row has smaller Mean Squared Error (Mean Absolute Error, Percentage Error) than the model of the column.

prediction problems. In those cases, the Stochastically Ordered Aggregation associated with the particular aggregation can be considered. It is also worth remarking that if the dimension of the aggregated vectors is too big, the optimization over the set of permutations can be unaffordable.

7.5 Tests for some stochastic orders

This section is devoted to briefly describe a possible application of the results presented in Sections 5.2 and 5.3, with some illustrative numerical examples. In particular, they will be used for testing stochastic orders between two distinct populations with unknown distributions, which is a common problem in many applicative fields.

Considering two random vectors $\vec{X}$ and $\vec{Y}$, some reasonable questions related to stochastic orders are whenever the inequalities $\vec{X} \leq_{cx} \vec{Y}$, $\vec{X} \leq_{sm} \vec{Y}$ and $X_1 \leq_{disp} Y_1$ under the conditions of Theorem 5.27 are true or not.

However, after searching in the literature, there are a few hypothesis testing procedures for the convex order, the supermodular order or the dispersive order between random vectors. Using the latter results, these tests can be constructed using as basis the well-known tests for the usual stochastic order or the increasing convex order, (see Section 2.4.3). In the following sections, the three cases will be discussed.

7.5.1 The convex order

Consider two simple random samples of size m of two random vectors $\vec{X}$ and $\vec{Y}$ of the same dimension n for which it is relevant to test the convex order. Then, the null hypothesis is determined as $H_0 : \vec{X} \leq_{cx} \vec{Y}$. Working under this hypothesis, Theorem 5.35 and Theorem 5.36 ensure that the quantities $\hat{\sigma}^2(\vec{X})$ and $\hat{\sigma}^2(\vec{Y})$, and the pair $G(\vec{X})$ and $G(\vec{Y})$, are ordered with respect to the increasing convex order. Then, it is possible to compute the values of such quantities in the sample and test for the increasing convex order among them. If the increasing convex order between the variability estimators is rejected, then the initial null hypothesis is also rejected.

A little numerical example has been developed to illustrate the latter method. Let U_1, U_2, U_3 and G be four independent random variables, having U_1, U_2 and U_3 standard uniform distribution and G standard Gaussian distribution. Consider the random vectors $\vec{X} = (U_1, U_2, U_3)$ and $\vec{Y} = (U_1 + G, U_2 + G, U_3)$. By using Theorem 2.117, it holds that $\vec{X} \leq_{cx} \vec{Y}$. 50 pairs of independent simple random samples of length 30 of $\vec{X}$ and $\vec{Y}$, $(\vec{X}_1, \ldots, \vec{X}_{30})$ and $(\vec{Y}_1, \ldots, \vec{Y}_{30})$ have been generated. Then, the associated sample values $(\hat{\sigma}^2(\vec{X}_1), \ldots, \hat{\sigma}^2(\vec{X}_{30}))$, $(\hat{\sigma}^2(\vec{Y}_1), \ldots, \hat{\sigma}^2(\vec{Y}_{30}))$ have been computed and the increasing convex order between them has been tested using the test given in [18]. The same has been done considering the pair $(G(\vec{X}_1), \ldots, G(\vec{X}_{30}))$ and $(G(\vec{Y}_1), \ldots, G(\vec{Y}_{30}))$.

The results can be found in Table 7.9. In all cases, it can be seen that the mean p-value is greater than the usual confidence level 0.05 and, therefore, there are no evidences to reject $H_0 : \vec{X} \leq_{cx} \vec{Y}$. However, if the null hypothesis is changed to $H_0 : \vec{Y} \leq_{cx} \vec{X}$, then the mean p-values are below 0.05. Notice that the results coincide with the order of the vectors, since by construction it is known that $\vec{X} \leq_{cx} \vec{Y}$. In addition, it seems that the sample variance leads to smaller p-values, at least in this

particular example.

Null hypothesis	Sample variance	Gini mean difference
$\vec{X} \leq_{cx} \vec{Y}$	0.8024	0.8064
$\vec{Y} \leq_{cx} \vec{X}$	0.0026	0

Table 7.9: Mean p-values associated with the test for the multivariate convex order when different null hypothesis and variability measures are considered.

7.5.2 The dispersive order

Consider two random vectors $\vec{X}$ and $\vec{Y}$ with identically distributed components ($X_1 =_{st} \cdots =_{st} X_n$ and $Y_1 =_{st} \ldots =_{st} Y_n$) with X_1 and X_2 fulfilling (P1) and (P2) as introduced in Section 5.2.1. That is,

(P1) Both X_1 and Y_1 have continuous distribution functions F and G,

(P2) The transformation $\phi = F^{-1} \circ G$ such that $X = \phi(Y)$ is strictly monotone,

where F is the distribution function of X_1 and G is the distribution function of Y_1. If the dispersive order wants to be tested between X_1 and Y_1, as a consequence of Theorems 5.29, and 5.30, one has that the sample values $\hat{\sigma}^2(\vec{X})$ and $\hat{\sigma}^2(\vec{Y})$, the sample values $G(\vec{X})$ and $G(\vec{Y})$ and the sample values $R(\vec{X})$ and $R(\vec{Y})$ are ordered with respect to the usual stochastic order.

Then, the null hypothesis $H_0 : X_1 \leq_{disp} Y_1$ can be tested by testing the usual stochastic order using the sample values of the variance, the Gini mean difference and the range. If the usual stochastic order between the variability estimators is rejected, then the initial null hypothesis is also rejected.

A similar experiment as in the last section has been considered. Consider the random vector $\vec{Y} = (U_1, U_2, U_3)$ with independent and uniformly distributed on $[0, 0.5]$ components. Consider, in addition, $\vec{X} = (U_1^2, U_2^2, U_3^2)$. Since the derivative of the square is smaller than 1 in $[0, 0.5]$, it is clear that it is a contraction on $[0, 0.5]$ and that $X_1 \leq_{disp} Y_1$. In addition, conditions (P1) and (P2) are fulfilled.

50 pairs of independent simple random samples of length 30 of $\vec{X}$ and $\vec{Y}$, $(\vec{X}_1, \ldots, \vec{X}_{30})$ and $(\vec{Y}_1, \ldots, \vec{Y}_{30})$, have been generated. Then, the associated sample

values $(\hat{\sigma}^2(\vec{X}_1),\ldots,\hat{\sigma}^2(\vec{X}_{30}))$, $(\hat{\sigma}^2(\vec{Y}_1),\ldots,\hat{\sigma}^2(\vec{Y}_{30}))$ have been computed and the usual stochastic order, using a modification of the Kolmogorov-Smirnov test given in the function *ks.test* in [274], has been tested. The same has been done considering the sample values $(G(\vec{X}_1),\ldots,G(\vec{X}_{30}))$ and $(G(\vec{Y}_1),\ldots,G(\vec{Y}_{30}))$, and the sample values $(R(\vec{X}_1),\ldots,R(\vec{X}_{30}))$ and $(R(\vec{Y}_1),\ldots,R(\vec{Y}_{30}))$.

Results can be found in Table 7.10. In all cases, it seems that the mean p-value is greater than the usual confidence level 0.05 and, therefore, there are is no evidence to reject $H_0 : X_1 \leq_{disp} Y_1$. However, if the null hypothesis is changed to $H_0 : Y_1 \leq_{disp} X_1$, then the mean p-values are below 0.05. Notice that the results coincide with the order of the vectors, since by construction $X_1 \leq_{disp} Y_1$.

Null hypothesis	Sample variance	Gini mean difference	Range
$X_1 \leq_{disp} Y_1$	0.9993	0.9991	0.9997
$Y_1 \leq_{disp} X_1$	0.0044	0.0082	0.0125

Table 7.10: Mean p-values associated with the test for the dispersive when different null hypothesis and variability measures are considered.

A similar procedure can be found in [303], in which the order between expectations of the Gini mean differences (which is a consequence of Corollary 5.28) is used to test some variability orders. However, this procedure was presented there only for the case of independent marginals. In addition, the usual stochastic order is a stronger condition than the order based on expectations, which allows one to construct more powerful tests.

7.5.3 The supermodular order

For the last case, consider two random vectors $\vec{X}$ and $\vec{Y}$ such that $\vec{X} \leq_{sm} \vec{Y}$. As a consequence of Corollary 5.41, the sample variance, Gini mean difference and the sample range are ordered with respect to the increasing convex order. Then, the null hypothesis $H_0 : \vec{X} \leq_{sm} \vec{Y}$ can be tested by testing the increasing stochastic order using the sample values of the latter measures. If the increasing convex order between them is rejected, then the supermodular order is rejected too.

Let $\vec{X}$ be a random vector with independent standard Gaussian components. In addition, consider $\vec{Y}$ a random vector with multivariate Gaussian distribution having a null mean vector and the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

By Theorem 11 in [247], $\vec{X} \leq_{sm} \vec{Y}$. Therefore, 50 pairs of independent simple random samples of length 30 of $\vec{X}$ and $\vec{Y}$, $(\vec{X}_1, \ldots, \vec{X}_{30})$ and $(\vec{Y}_1, \ldots, \vec{Y}_{30})$, have been generated. Then, the associated sample values $(\hat{\sigma}^2(\vec{X}_1), \ldots, \hat{\sigma}^2(\vec{X}_{30}))$ and $(\hat{\sigma}^2(\vec{Y}_1), \ldots, \hat{\sigma}^2(\vec{Y}_{30}))$ have been computed and the increasing convex order between them has been tested using the test given in [18]. The same has been done considering the sample values $(G(\vec{X}_1), \ldots, G(\vec{X}_{30}))$ and $(G(\vec{Y}_1), \ldots, G(\vec{Y}_{30}))$ and the sample values $(R(\vec{X}_1), \ldots, R(\vec{X}_{30}))$ and $(R(\vec{Y}_1), \ldots, R(\vec{Y}_{30}))$.

The results can be found in Table 7.11. In all cases, the mean p-value is greater than the usual confidence level 0.05 and, therefore, there is no evidence to reject $H_0 : \vec{X} \leq_{sm} \vec{Y}$. However, if the null hypothesis is changed to $H_0 : \vec{Y} \leq_{sm} \vec{X}$, then the mean p-values are below 0.05. Notice that the results coincide with the order of the vectors, since by construction $\vec{X} \leq_{sm} \vec{Y}$.

Null hypothesis	Sample variance	Gini mean difference	Range
$\vec{X} \leq_{sm} \vec{Y}$	0.7602	0.7953	0.8002
$\vec{Y} \leq_{sm} \vec{X}$	0.0262	0.0366	0.0430

Table 7.11: Mean p-values associated with the test for the supermodular order when different null hypothesis and variability measures are considered [40].

It should be noted that a similar approach is given in [89], using the fact that, if $\vec{X} \leq_{sm} \vec{Y}$, then $\max_{i \in I} X_i \geq_{st} \max_{i \in I} Y_i$ and $\min_{i \in I} X_i \leq_{st} \min_{i \in I} Y_i$ for any $I \subseteq [n]$, see [24].

7.6 Time series forecasting using the IOLF operator

Time series forecasting is a task used to obtain estimates of future values of various measurements in the real world. Models used for predicting time series values are commonly regressors, i.e. any variable in a regression model that is used to predict a response variable. If a set of regression models is considered, they may have several problems: the first is that the models are weak when taken individually (there are models that obtain admissible errors only in some cases) and the second is that the regressors have similar effects. In order to try to avoid these two problems and obtain the most optimal model, combining them becomes a fundamental task.

As discussed in Section 4.1.1, aggregation functions are often used as ensembles that fuse the prediction of different prediction models. This section is devoted to illustrate the benefits of the IOLF operator, see Definition 4.1, in this regard. In particular, Weighted arithmetic means, OWA operators and IOWA operators, as well as their counterparts with negative weights are used in a problem related to the forecasting of temperature and humidity prediction.

7.6.1 Description of the experimental procedure

In this example, forecasts of different prediction models are fused by means of aggregation and pre-aggregation functions. Three of the ensemble alternatives are the Weighted Arithmetic Mean (WAM), the Ordered Weighted Averaging (OWA) and the Induced Ordered Weighted Averaging (IOWA) operators considering as the inducing vector the precision in the previous time step, as explained in Section 4.1.2. These alternatives are the most used among aggregation functions in the literature, see Table 4.1. The other three alternatives are the Linear Fusion (LF), the Ordered Linear Fusion (OLF) and the Induced Ordered Linear Fusion (IOLF) operator, which are constructed exactly as, respectively, WAM, the OWA and the IOWA operators but allowing the weights to be negative. Notice that the first three cases can be seen as IOWA operators with different inducing vectors and the other cases as the same models but allowing negative weights, which are all particular cases of the IOLF operator introduced in Definition 4.1. In this example, 7 different prediction models, which will be explained later, are considered.

Taking into account the notation of Section 4.1.2, $p_{i,t}$ is the prediction of the model i at time t, and therefore it has to be considered the vector $\vec{p} = (p_{1,t}, \ldots, p_{7,t})$ of the predictions of the 7 models at time t. Then, use Equation 4.2 to obtain the vector $\vec{y} = (y_{1,t-1}, \ldots, y_{7,t-1})$. This vector will be used as the vector that induces the order in the IOWA and IOLF cases. In the cases of OWA and OLF operators, predictions are ordered from the smallest to the greatest for each time. Finally, no additional ordering is made for the WAM and LF.

Therefore, the fused predicted values (for a time t) are the ones obtained in the following way:

$$\hat{p}_{IOWA} = \text{IOWA}(\vec{p}, \vec{y}; \vec{w}) = \sum_{i=1}^{7} w_i \pi_{\vec{y}}(\vec{p})_{i,t}, \qquad \hat{p}_{IOLF} = \text{IOLF}(\vec{p}, \vec{y}; \vec{w}),$$

$$\hat{p}_{OWA} = \text{OWA}(\vec{p}; \vec{w}) = \sum_{i=1}^{7} w_i p_{\sigma_t(i),t}, \qquad \hat{p}_{OLF} = \text{OLF}(\vec{p}; \vec{w}),$$

$$\hat{p}_{WAM} = \text{WAM}(\vec{p}; \vec{w}) = \sum_{i=1}^{7} w_i p_{i,t}, \qquad \hat{p}_{LF} = \text{LF}(\vec{p}; \vec{w}),$$

where σ_t is a permutation such that $p_{\sigma_t(1),t} \geq \cdots \geq p_{\sigma_t(7),t}$.

The optimal weights for IOWA, OWA and WAM operators are computed numerically by solving the problem stated in Equation 4.3. In the case of the IOLF, OLF and LF operators, the optimal weights can be computed directly by applying Theorem 4.6. It should be noted that, in some cases, inverting the matrix $\Sigma + \vec{\Delta}\vec{\Delta}^t$ cannot be easy if the dimension is too high or it is ill-conditioned. In these cases, the optimal weights can be computed numerically, as done in the cases of the IOWA, OWA and WAW operators.

The considered data, consisting of the temperature and humidity of different places in a house, have been obtained from [79] and is composed of almost 20000 observations, which are measured every 10 minutes for about 4.5 months. The house temperature and humidity conditions are monitored with a ZigBee wireless sensor network. The time series measure the temperature (T) and humidity (RH) in 9 different areas. These time series have several characteristics in common. Firstly, there exists a strong seasonal component with a period of one day. On the other hand, there does not exist a weekly or monthly seasonal component. Secondly, a moderate amount of outliers appear in the data. Thirdly, the values of the time

series seem to have a bell-shaped distribution. For more information in this regard, the reader is referred to [80].

For each of these 18 time series, the data were divided into the first 70% of the days, the training sample, and the remaining 30%, the test sample. On the training sample, seven different forecasting models were fitted. The used models are the following:

RF Random Forest [160]. This regression method builds a set of decision trees in the training process. It returns the average prediction of the individual trees. In this case, the number of trees (estimators) to be used is set to 1000.

GB Gradient Boosting [132]. This method uses weak decision trees, which are boosted by the gradient. A gradient boosting model is built in stages by optimizing based on a cost function. The number of stages is set to 1000. The learning rate of the model is set to 0.1 and the loss function used is the Friedman Mean Squared Error [131].

ARI ARIMA [68]. Autoregressive integrated moving average (ARIMA) model is a statistical model that uses variations and regressions of statistical data to find patterns for future prediction. It is a model that tries to identify coefficients and the number of regressions to be used and, since it is a dynamic model, predictions are not based on independent variables but on past data. In this case, the parameters are fitted according to the Akaike's Information Criterium.

KNN Regression based on K-Nearest Neighbors [98] with $k = 3$. This classical method is based on the entry of the k nearest values in the dataset. The predicted value is assigned to the average of the values of the k nearest neighbors. The function used to measure the distance between values has been the Euclidean distance.

BR Bagging regressor [71]. Bagging is a general variance reduction method based on the use of bootstrap together with a decision tree. For regression trees, many trees are grown (without pruning) and the mean of the predictions is calculated. An additional advantage of bagging is that it allows one

to estimate the prediction error directly, without the need to use a test sample or to apply cross-validation. The number of trees used in this model is fixed to 1000.

Finally, models 6 and 7 are models based on recurrent neural networks (RNN). Specifically, models based on Long Short-Term Memories (LSTM) [161] and Gated Recurrent Unit (GRU) [88] neurons are considered. These models, introduced in 1997 and 2014 respectively, were designed to avoid the training problems of classical RNNs with input data having long dependencies. To avoid these problems, these neural networks use multiplicative units, called gates, which control the flow of information that is fed into the network. On the other hand, these models allow information to be stored in short and long term memories for use at future points in time.

LSTM LSTM-based model [161]. The first RNN-based model consists of a single LSTM layer with a hidden size of 64 elements and a dense layer. The loss function used in this model is the Mean Squared Error, and the optimization method is the Adam algorithm with a learning rate of 0.01. 1200 epochs have been run. At the input of the LSTMs the data are normalized to the range (-1,1) by min-max, and denormalized at the output.

GRU GRU-based model [88]. The second RNN-based model model is composed of a single GRU layer with a hidden size of 128 elements and a dense layer. A dropout of 0.4 has been used. The loss function used in this model is the Mean Squared Error, and the optimization method is the Adam algorithm with a learning rate of 0.01. 1200 epochs have been run. The input data are normalized to the range (-1,1) by min-max, and denormalized at the output.

Notice that the latter models belong to different families and their behavior will be, in general, very different. The procedure can be summarized in 4 steps, as can be seen in Figure 7.2:

1. Train the prediction models,

2. Reorder the prediction vectors for each time,

3. Compute the optimal weights,

4. Obtain the fused predictions.

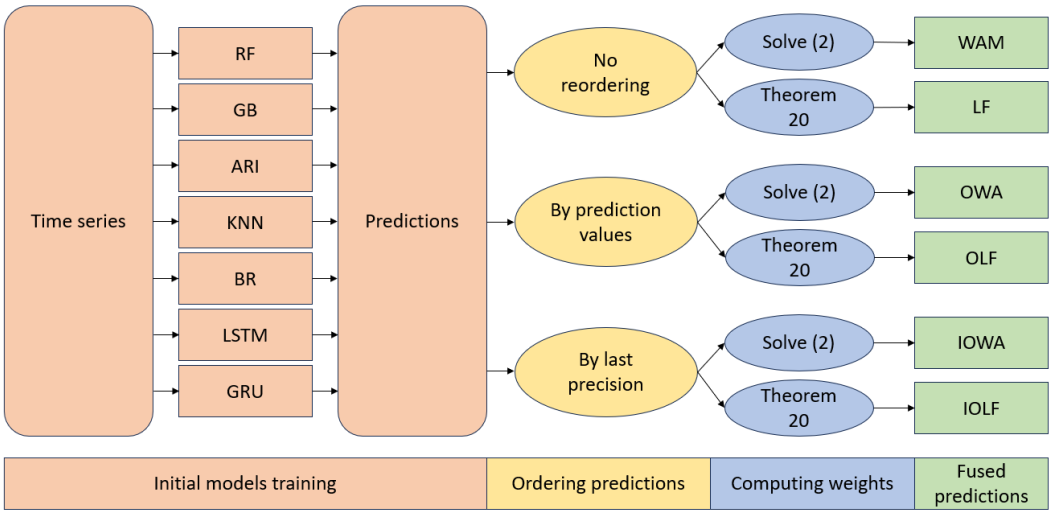

Figure 7.2: Diagram showing the followed steps to obtain the fused prediction of each considered ensemble [35].

7.6.2 Discussion of the results

The performance of the initial models and the fusion operators is compared using the test sample. Since some of the forecasting models involve random initialization or dropout, the procedure has been replicated ten times for each of the time series. The result for each time series and prediction model and fusion operator in the test sample can be found in Table 7.13. For the Mean Squared Error, in four of the time series the best performance was made by an initial prediction model, in other four one of the classical aggregation functions was the best option and in the remaining ten time series, one of the proposed fusion operators with negative weights had the lowest MSE.

In order to compare globally the performance of the different alternatives, the boxplot associated with the relative error of all executions, ten for each of the eighteen time series, can be found in Figure 7.3. The relative error has been computed

by dividing the square of the MSE by the mean value of the corresponding time series.

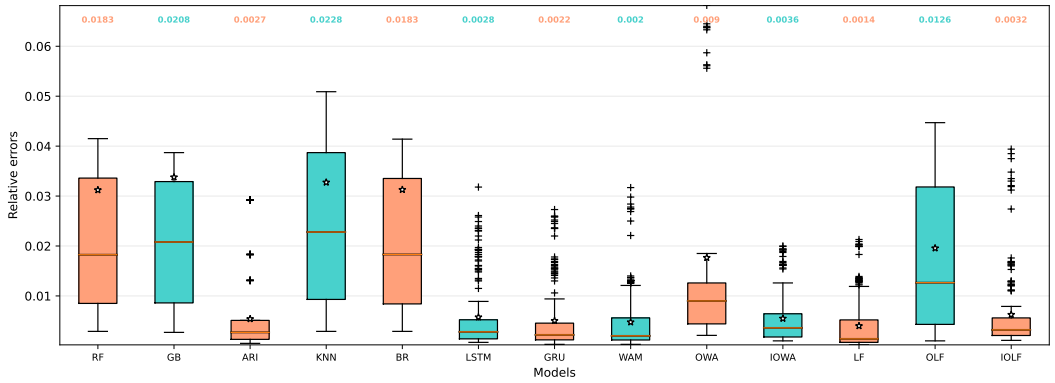

Figure 7.3: Boxplot of the relative errors of the different models and their aggregations. The median the errors for each model and aggregation is at the top of the graph [35].

Qualitatively, the best initial prediction models seem to be ARIMA, LSTM and GRU. Between the fusion operators, the WAM and LF seem to be the better options, followed by the IOWA and IOLF. In order to have a quantitative comparison, Wilcoxon tests (see Section 2.4.3) have been performed, considering the 180 paired samples of Mean Squared Errors. The p-values can be found in Table 7.12, in which the alternative hypothesis has been chosen as the row model having a smaller MSE than the column model.

As can be seen, it is statistically accepted that the Linear Fusion operator is better than the rest of alternatives. In addition, the p-value between the IOLF and the IOWA operator is 0.112, which is small but not enough to be able to draw a conclusion. Nothing can be said about the comparison between the OWA and OLF operators. In order to facilitate the visualization of the hierarchy between the alternatives, the thirteen alternatives are ordered from right to left, the ones on the left being significantly better than the others (see Figure 7.4).

Notice that the GRU model is the best among all the initial prediction models and is also better than OWA, IOWA, OLF and IOLF. This can be explained because, since these fusion operators exchange the prediction positions for each time, it is

	RF	GB	ARIMA	KNN	BR	LSTM	GRU	WAM	OWA	IOWA	LF	OLF	IOLF
RF	-	0.017	1	0	0.536	1	1	1	1	1	1	1	1
GB	0.983	-	1	0	0.983	1	1	1	1	1	1	1	1
ARIMA	0	0	-	0	0	0.752	1	1	0	0.744	1	0	0.388
KNN	1	1	1	-	1	1	1	1	1	1	1	1	1
BR	0.464	0.017	1	0	-	1	1	1	1	1	1	1	1
LSTM	0	0	0.248	0	0	-	0.997	1	0	0.431	1	0	0.086
GRU	0	0	0	0	0	0	-	0.997	0	0	1	0	0
WAM	0	0	0	0	0	0	0.003	-	0	0	1	0	0
OWA	0	0	1	0	0	1	1	1	-	1	1	0.360	1
IOWA	0	0	0.257	0	0	0.569	1	1	0	-	1	0	0.888
LF	0	0	0	0	0	0	0	0	0	0	-	0	0
OLF	0	0	1	0	0	1	1	1	0.640	1	1	-	1
IOLF	0	0	0.612	0	0	0.914	1	1	0	0.112	1	0	-

Table 7.12: P-values associated with the Wilcoxon test with null alternative hypothesis: the row model has a bigger Mean Squared Error than the column model [35].

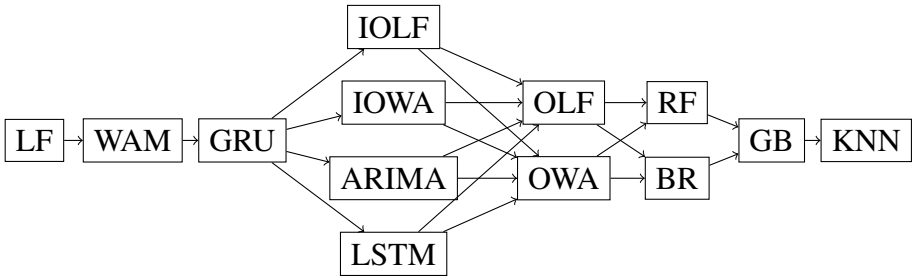

Figure 7.4: Prediction models and fusion operators ordered from left to right according to having a significantly smaller MSE according to the Wilcoxon test [35].

hard for them to give importance to a particular model.

It is important to remark that the benefits of the here-presented methods are related to the appearance of negative weights in the optimization process. This is the case of the Linear Fusion in the latter example, which outperforms the rest of alternatives. However, there are databases and elections of forecasting models in which the negative weights do not appear or are not relevant. In these cases, the behavior allowing or not negative weights should be similar. This could be the case for the OWA and OLF operators.

						Mean Squared Error						
DS	RF	GB	ARI	KNN	BR	LSTM	GRU	WAM	OWA	IOWA	LF	OLF	IOLF
T1	0.256±0.00105	0.276±0.00264	0.000423±0	0.271±3e−05	0.255±0.000708	0.000849±0.00130	0.000323±0.000146	0.00121±0.00058	0.0566±0.00657	0.0220±0.0182	**0.000236±0.00017**	0.0936±0.00638	0.0214±0.0178
RH1	0.188±0.000976	0.219±0.00141	0.0430±0	0.436±4e−06	0.188±0.00144	0.0419±0.00442	0.0417±0.00403	0.0506±0.00517	0.132±0.00371	0.0499±0.00246	0.0499±0.00992	0.0942±0.00227	**0.0439±0.00156**
T2	0.602±0.00105	0.618±4.8e−05	0.00863±0	0.619±3e−06	0.602±0.00102	**0.00638±0.00131**	0.00808±0.00116	0.0305±0.0108	0.0505±0.0106	0.0101±0.00384	0.0111±0.00313	0.599±0.00205	0.00698±0.00179
RH2	0.765±0.00422	0.953±0.00548	0.0425±0	1.00±3e−06	0.766±0.00296	0.215±0.478	0.0153±0.00284	0.0112±0.00335	0.384±0.00622	0.0248±0.0102	**0.00802±0.00271**	0.267±0.0256	0.0205±0.00779
T3	0.0794±0.000328	0.0833±0.000625	0.00130±0	0.116±4e−06	0.0794±0.00023	0.00183±0.00379	0.00104±0.000435	**0.00725±0.000641**	0.0163±0.00303	0.000918±0.000588	0.000959±0.00821	0.0513±0.0057	0.000976±0.00031
RH3	0.0516±0.000106	0.0452±0.000304	0.0122±0	0.0628±8e−06	0.0517±0.000234	0.00434±0.00186	0.00409±0.000824	0.00315±0.00142	0.0291±0.00103	0.00454±0.000801	**0.00290±0.00127**	0.0129±0.00087	0.00473±0.00703
T4	0.496±0.00287	0.471±5.7e−05	**0.00737±0**	0.478±9e−06	0.489±0.00249	0.00223±0.00516	0.0140±0.0389	0.00229±0.00237	0.0334±0.0107	0.00111±0.00113	0.0011±0.00145	0.426±0.0461	0.00458±0.00274
RH4	0.0312±0.000112	0.0321±0.000118	0.00777±0	0.0392±4e−06	0.0312±0.000136	0.0457±0.115	0.00147±0.000334	0.000661±0.000542	0.0166±0.00877	0.00325±0.000673	**0.000589±0.00495**	0.00761±0.000287	0.00321±0.00065
T5	0.655±0.00335	0.424±0.00541	**0.00769±0**	0.949±1.2e−05	0.649±0.00455	0.00311±0.00145	0.00437±0.00133	0.00641±0.00386	0.00709±0.00568	0.0115±0.00564	0.00636±0.00792	0.497±0.0394	0.00575±0.00403
RH5	16.0±0.61	45.8±0.116	0.448±0	6.72±2e−06	16.6±0.805	0.747±0.208	0.672±0.0923	**0.444±0.00743**	9.83±1.31	0.682±0.0361	0.448±0.0209	4.16±0.405	0.705±0.0494
RH6	0.661±0.00127	0.914±0.00604	0.997±0	1.41±0	0.661±0.00241	0.914±0.214	0.773±0.497	0.437±0.119	0.505±0.024	**0.407±0.0473**	0.426±0.109	0.476±0.027	0.439±0.0525
T7	0.268±0.000603	0.277±3e−06	0.000196±0	0.258±7e−06	0.268±0.000523	0.000479±0.000293	0.000127±0.000121	0.000312±0.000448	0.0547±0.00951	0.00873±0.00755	**0.000118±9.4e−05**	0.106±0.00787	0.00815±0.0076
RH7	0.0105±2.7e−05	0.00915±2.6e−05	0.00854±0	0.0107±3e−06	0.0105±4.5e−05	0.0529±0.145	0.00299±0.000512	0.00156±0.00103	0.00566±0.000405	0.00312±0.000576	**0.00139±0.000954**	0.00357±0.00123	0.00318±0.00094
T8	0.105±0.000305	0.101±0.000242	0.00523±0	0.114±3.3e−05	0.106±0.000294	0.00152±0.00162	0.000142±3.9e−05	0.000232±0.000183	0.0339±0.00693	0.00717±0.00516	**9.71e−05±9.42e−05**	0.0368±0.00649	0.00696±0.00516
RH8	0.0219±2.6e−05	0.0208±1.8e−05	0.0242±0	0.0221±1e−06	0.0219±2.3e−05	0.0129±0.00191	0.0163±0.00796	0.0108±0.00144	0.0157±0.000496	0.0123±0.00107	**0.0101±0.00198**	0.0137±0.000537	0.0126±0.00114
T9	0.321±0.00647	0.233±0.00135	**8.56e−05±0**	0.905±4e−06	0.323±0.00729	0.0126±0.0125	0.00268±0.00213	0.00125±0.000633	0.0585±0.000866	0.0148±0.0039	0.00238±0.000155	0.185±0.00235	0.0123±0.0038
RH9	0.123±0.000821	0.128±0.00106	0.0142±0	0.153±6e−06	0.123±0.000681	0.0144±0.008	0.0224±0.00438	0.00432±0.00198	0.0714±0.00762	0.00864±0.00476	**0.00333±0.00196**	0.0327±0.00253	0.00844±0.00417
						Mean Absolute Error							
DS	RF	GB	ARI	KNN	BR	LSTM	GRU	WAM	OWA	IOWA	LF	OLF	IOLF
T1	0.258±0.000673	0.268±0.00158	0.0149±0	0.277±0.00043	0.257±0.000436	0.0179±0.0132	0.0134±0.00332	0.0210±0.00491	0.125±0.00669	0.0755±0.0251	**0.0107±0.00318**	0.157±0.00535	0.0746±0.0251
RH1	0.183±0.000372	0.183±0.000341	0.0981±0	0.261±8.2e−05	0.183±0.000459	0.0998±0.00418	0.103±0.0169	0.107±0.00577	0.135±0.00169	0.102±0.00246	0.113±0.0105	0.137±0.00116	**0.0953±0.002**
T2	0.291±0.000362	0.296±4.8e−05	0.0526±0	0.308±3.5e−05	0.292±0.000369	**0.0466±0.00426**	0.0561±0.00253	0.0872±0.0158	0.108±0.00949	0.0592±0.00962	0.0644±0.00795	0.293±0.00417	0.0532±0.00459
RH2	0.308±0.00102	0.333±0.0114	0.110±0	0.361±1.5e−05	0.309±0.000542	0.219±0.262	0.0741±0.00823	0.0539±0.00698	0.217±0.0154	0.0791±0.0118	**0.0503±0.008**	0.186±0.0149	0.0743±0.011
T3	0.092±0.000447	0.0909±0.000801	0.0198±0	0.133±0.000106	0.0920±0.000261	0.0237±0.0253	0.0205±0.00457	**0.0154±0.0083**	0.0454±0.00268	0.0165±0.00344	0.0189±0.00953	0.0806±0.00483	0.0185±0.00252
RH3	0.0802±6.6e−05	0.0715±0.000108	0.0597±0	0.0909±0.000186	0.0802±0.000127	0.0418±0.0122	0.0358±0.00294	**0.0341±0.00791**	0.0563±0.00985	0.0374±0.00265	0.0344±0.00826	0.0513±0.00149	0.0383±0.00239
T4	0.362±0.00139	0.348±5.8e−05	**0.0161±0**	0.361±0.000124	0.362±0.00122	0.0271±0.0329	0.0570±0.0981	0.0306±0.0151	0.0964±0.0161	0.0215±0.008	0.0219±0.0162	0.339±0.0299	0.0401±0.00878
RH4	0.0743±8.9e−05	0.0682±7.8e−05	0.0496±0	0.0836±8.8e−05	0.0743±9.1e−05	0.105±0.17	0.0241±0.00257	0.0153±0.00559	0.0486±0.00183	0.0321±0.00359	**0.0150±0.00543**	0.0385±0.00554	0.0322±0.00362
T5	0.449±0.0013	0.349±0.00269	**0.0110±0.0**	0.565±0.000229	0.447±0.00173	0.0363±0.00812	0.0449±0.00841	0.049±0.0152	0.0505±0.0122	0.0619±0.0154	0.0376±0.0313	0.392±0.0175	0.0388±0.017
RH5	0.903±0.0108	1.32±0.00165	0.145±0	0.703±1.9e−05	0.913±0.0144	0.414±0.141	0.377±0.0605	**0.168±0.00951**	0.692±0.0275	0.239±0.00674	0.186±0.0271	0.551±0.0227	0.243±0.0144
T6	0.597±0.00101	0.595±0.00127	0.163±0	0.639±2.2e−05	0.597±0.000716	0.127±0.0209	0.135±0.00689	0.138±0.038	0.423±0.0164	**0.107±0.014**	0.112±0.019	0.315±0.023	0.165±0.0152
RH6	0.532±0.000844	0.573±0.00125	0.547±0	0.681±0	0.532±0.000737	0.556±0.0685	0.576±0.283	0.429±0.0602	0.471±0.00971	**0.381±0.0277**	0.421±0.0699	0.466±0.0177	0.381±0.0296
T7	0.248±0.000338	0.25±3.8e−05	0.00891±0	0.252±0.000154	0.248±0.000402	0.015±0.00459	0.00728±0.00331	0.00901±0.00418	0.112±0.00904	0.0441±0.0168	**0.00658±0.00263**	0.158±0.00626	0.0422±0.0175
RH7	0.0719±6.5e−05	0.0687±4.3e−05	0.0611±0	0.0743±8e−06	0.0719±8.6e−05	0.109±0.195	0.0374±0.00326	0.0268±0.0106	0.0534±0.00194	0.0382±0.00405	**0.0255±0.0101**	0.0434±0.00915	0.0381±0.00409
T8	0.127±0.000235	0.120±0.000179	0.0175±0	0.137±0.000624	0.128±0.000221	0.0265±0.0149	0.00837±0.000861	0.00857±0.00217	0.0738±0.00614	0.0353±0.00947	**0.00606±0.00252**	0.0750±0.0063	0.0348±0.00965
RH8	0.111±7.6e−05	0.108±3.3e−05	0.109±0	0.112±1e−05	0.111±4.7e−05	0.0811±0.0063	0.0929±0.0284	0.0762±0.00561	0.0933±0.00153	0.0800±0.00401	**0.0739±0.00822**	0.0897±0.00163	0.0808±0.00424
T9	0.363±0.00389	0.284±0.00169	**0.00529±0**	0.635±0.00016	0.362±0.00384	0.0796±0.0426	0.0382±0.0173	0.0236±0.00592	0.151±0.0106	0.0789±0.0116	0.0102±0.0039	0.279±0.0185	0.0710±0.0119
RH9	0.135±0.00044	0.134±0.000408	0.0801±0	0.153±2.9e−05	0.135±0.000319	0.0737±0.0242	0.0919±0.0921	0.0394±0.0123	0.106±0.00427	0.0551±0.0083	**0.0360±0.0125**	0.0767±0.00957	0.0546±0.00797

Table 7.13: Mean and Standard Deviation (SD) of Mean Squared Error and Mean Absolute Error for the prediction models and their aggregations in the considered time series. The lowest error of each row is in bold [35].

Conclusions

In this thesis, extensions of the concept of aggregation functions for random structures have been defined and studied in detail.

First, the simplest case, random variables, has been considered. In order to define such a concept, a stochastic order has to be considered, concluding that the appropriate choice is the usual stochastic order. Some families of aggregations of random variables have been defined, showing that many considerations that cannot be done with usual aggregation functions, such as the identification of the input random vector, the introduction of random parameters or the changes in the dependence between inputs and output, can be done with aggregations of random variables. Moreover, it has been seen that idempotence and internality are not straightforward notions for aggregations of random variables and admit different extensions.

Some specific examples of aggregations of random variables are then given. Working with the Induced Ordered Linear Fusion (IOLF) operator, it is shown that considering negative weights, besides extending of the feasible region in optimization problems, gives a closed expression of the optimal weights, which is useful for proving properties of the operator as an estimator. Then, a special case of the IOLF operator is used to prove an asymptotic result considering the convergence of the optimal weights for mean estimation of symmetric distributions by using order statistics.

The idea of ordering distributions rather than values is explored by defining the Stochastically Ordered Aggregations, which perform a pointwise ordering of the involved distribution functions before applying an usual aggregation function. Unfortunately, the resulting functions are not monotone with respect to the usual stochastic order, but they are monotone with respect to a modification that requires

the same dependence structure.

The case of uniform random capacities is also considered. Using the probability approach, many properties of them have been proved, showing that they have a deep relation with the order statistics of independent standard random variables. In addition, an open problem in the literature related to the minimization of the distance between two random variables is solved in its more general scenario. Its solution can be applied to define aggregations of random variables.

In the thesis, contributions to the theory of stochastic orders can also be found. In particular, it is concluded that some families of mean functions reduce the variability (with respect to the convex stochastic order) when applied to random vectors. It is also concluded that many variability measures take greater values when the associated random vector has more variability or less positive dependence. Although these results are intuitive, formal results in this regard cannot be found in the literature.

Then, the extension of the notion of aggregation of random variables to other random structures is studied, with the cases of random elements on bounded posets, random vectors, stochastic orders and random sets being the ones studied more in detail. In addition to the construction of such notions, it is concluded that many properties of random vectors and random processes are preserved when the same usual aggregation function is applied componentwise or pointwise, respectively. For random sets, a proper location stochastic order has been defined, a concept that, to the best knowledge of the author, was not previously considered in the literature.

Finally, some of the methods presented have been applied to problems involving real or simulated data. In all cases, it has been concluded that the methods worked as intended. Moreover, in the cases where a new aggregation operator has been compared with usual aggregation functions for a prediction task, the proposal has been significantly better. However, it has to be pointed out that the experimental results can differ if other databases are considered. In addition, it seems difficult to have aggregations of random variables to outperform more involved models, such as neural networks, in prediction problems. Of course, aggregations of random variables can have better semantics, but for applied problems the best could be to use them in combination with machine learning models, as in Section 7.6, rather by their own.

All theoretical and experimental results lead to the three main conclusions of the thesis.

1. **The aggregation of random structures is feasible.** Working with random elements instead of elements includes many technical issues in the construction of the notion of aggregation such as measurability, conditional distributions or the existence of some necessary sets of random elements. However, as proved in Chapter 3 and Chapter 6, such a construction is possible.

2. **The aggregation of random structures extends usual Aggregation Theory.** Firstly, the possibility of the use of usual aggregation functions as aggregations of random structures is stated in the several Composition Theorems. In addition, it has been proven that there are many cases of aggregations of random structures that cannot be reproduced by considering usual aggregation functions, as seen in Section 3.2.

3. **The aggregation of random structures is applicable to data.** Leaving aside theoretical properties, the introduced methods can be computed and applied to real data, as illustrated in Chapter 7. In addition, the applicability of results involving aggregation of random structures, since many different branches of Mathematics are involved, can trespass Aggregation Theory, see Section 7.5.

Therefore, it is possible to state that the initial fixed objectives, which were 1. define and study the concept of aggregation for different random structures, 2. define new aggregation operators, study their properties and use them in applied problems and 3. extend the behavioral study of aggregation functions by means of random variables, have been attained.

However, some questions are still without an answer. Leaving aside Section 6.5, in which just brief considerations have been considered and a deep study should be done in the future, the main open problems are the following. (1) Is there a meaningful formula that can describe any aggregation of random variables in terms of a conditionally determined and a randomly induced aggregations of random variables? (2) Does the convergence of the cumulative optimal weights also hold for non-symmetric distributions? (3) Is it possible to derive the exact distribution of the

measure of a particular subset of a uniform random capacity? (4) Is there a solution to the problem in Section 4.5 when considering random vectors? (5) What are the conditions for the usual stochastic ordering of two variability measures when considering dependence stochastic orders between the initial random vectors? and (6) Is it possible to define a proper location stochastic order for random sets that is closed under intersection, union and the Cartesian product?

Leaving aside the open questions, this work is intended to be a first step in the use of random structures in Aggregation Theory when applied to data analysis. In this regard, the reader is recalled again, as explained in Section 1.3, about the importance of the use of probability and Statistics procedures in Aggregation Theory. Probability Theory in aggregation is not only appropriate but indispensable.

Conclusiones

En esta tesis se ha definido y estudiado el concepto de agregación de estructuras aleatorias.

En primer lugar, se ha considerado el caso de variables aleatorias, el más sencillo. En la definición de dicho concepto, es necesario elegir un orden estocástico, siendo la opción más apropiada el orden estocástico usual. Después, se han definido algunas familias de agregaciones de variables aleatorias que permiten considerar escenarios que no son posibles con agregaciones usuales, como la identificación del vector aleatorio que se está agregando, la introducción de parámetros aleatorios o la posibilidad de cambios en la dependencia entre las variables aleatorias agregadas y el resultado de la agregación. Además, se ha visto que no es inmediato extender el concepto de internalidad e idempotencia a agregaciones de variables aleatorias, ya que hay más de una alternativa.

Posteriormente, se ha trabajado con tipos de agregaciones de variables aleatorias especialmente relevantes. Considerando el operador de fusión lineal ordenado e inducido (IOLF por sus siglas en inglés), se ha demostrado que, además de permitir ampliar la región factible en algunos problemas de optimización, considerar pesos negativos permite dar una expresión cerrada de los pesos óptimos, lo que es útil a la hora de demostrar propiedades de dicho operador como estimador. Adicionalmente, se ha utilizado un caso particular del operador IOLF para demostrar la convergencia de los pesos óptimos acumulados en problemas de estimación de la media cuando se consideran estadísticos ordenados de distribuciones simétricas.

Se han definido las agregaciones estocásticamente ordenadas, basándose en la idea de ordenar distribuciones en vez de los valores de la muestra. En particular, se definen mediante la ordenación punto a punto de las funciones de distribución

involucradas, para después aplicar una agregación usual. Desafortunadamente, estas agregaciones no son monótonas respecto al orden estocástico usual, aunque sí lo son si se realiza una modificación del mismo que requiere que los vectores aleatorios tengan la misma estructura de dependencia.

Adicionalmente, se ha hecho un estudio de capacidades aleatorias uniformes. Haciendo uso del enfoque probabilista, se han demostrado diferentes propiedades de este tipo de distribuciones, mostrando que tienen una relación estrecha con los estadísticos ordenados de variables aleatorias uniformes, estándar e independientes. Por otro lado, se ha considerado el problema de minimización de la distancia entre dos variables aleatorias cuando una de ellas y la dependencia entre las dos están fijadas en su forma más general, un problema que llevaba abierto hasta la fecha. La solución de dicho problema permite definir algunas agregaciones de variables aleatorias.

Esta tesis también incluye contribuciones a la rama de ordenaciones estocásticas. En particular, se dan resultados sobre la reducción de la variabilidad (respecto al orden estocástico convexo) que sucede al aplicar algunas funciones promedio. Se concluye también que, para muchas medidas de variabilidad, estas toman valores más grandes cuando el vector aleatorio asociado es más variable o tiene una dependencia negativa más fuerte. Resultados de este tipo, aunque intuitivos, no se pueden encontrar demostrados formalmente en la literatura.

El concepto de agregación de variables aleatorias se ha extendido a otras estructuras aleatorias, dedicando más atención al caso de elementos aleatorios sobre conjuntos parcialmente ordenados y acotados, vectores aleatorios, procesos estocásticos y conjuntos aleatorios. Además de la construcción de dichos conceptos, se ha concluido que varias propiedades de vectores aleatorios o procesos estocásticos se preservan al aplicar la misma función de agregación a, respectivamente, cada una de las componentes o de índices asociados. Respecto a los conjuntos aleatorios, se ha tenido que definir un orden estocástico de localización para los mismos, concepto que no se había considerado hasta la fecha.

Finalmente, algunos de los métodos desarrollados se han aplicado a problemas que involucran datos, ya sean reales o simulados. En todos los casos, el comportamiento de los mismos ha sido el esperado. Adicionalmente, en los casos en los que un nuevo operador de agregación se ha comparado con funciones de agre-

gación usuales, la nueva propuesta ha tenido un comportamiento significativamente mejor. Sin embargo, es posible que estos resultados experimentales difieran si se consideran otras bases de datos diferentes. De hecho, parece complicado que alguna agregación de variables aleatorias pueda mejorar los resultados de modelos más complejos como las redes neuronales, al menos en problemas de predicción. Si bien es cierto que las agregaciones de variables aleatorias pueden tener mayor interpretabilidad, para este tipo de problemas lo mejor seguramente sea combinarlas con modelos de aprendizaje automático, tal como se ha hecho en la sección 7.6.

Teniendo en cuenta todos los desarrollos y los resultados, ya sean teóricos o experimentales, se puede llegar a las tres conclusiones principales.

1. **Se pueden definir agregaciones de variables aleatorias.** Considerar elementos aleatorios añade cuestiones técnicas como la medibilidad, la existencia de ciertas colecciones de elementos aleatorios o distribuciones condicionadas. Sin embargo, tal como se ha expuesto en los capítulos 3 y 6, dichas construcciones son posibles.

2. **La agregación de estructuras aleatorias extiende la teoría de agregación usual.** En primer lugar, la capacidad de usar funciones de agregación usuales para definir agregaciones de estructuras aleatorias se ha probado mediando teoremas de composición. Además, se ha probado que hay muchos casos en los cuales las agregaciones de variables aleatorias no pueden ser reproducidas considerando agregaciones usuales, tal como se ha demostrado en la sección 3.2.

3. **Es posible aplicar agregaciones de estructuras aleatorias a datos.** Dejando de lado propiedades teóricas, los métodos que han sido introducidos pueden ser aplicados a datos reales, tal como se ha ilustrado en el capítulo 7. Adicionalmente, ya que diferentes ramas de las matemáticas están involucradas en el desarrollo teórico, pueden aparecer aplicaciones no directamente relacionadas con agregaciones, como por ejemplo la expuesta en la sección 7.5.

Por lo tanto, se puede afirmar que los objetivos inicialmente fijados, que eran 1. definir y estudiar el concepto de agregación de diferentes estructuras aleatorias,

2. definir nuevos operadores de agregación, estudiar sus propiedades y utilizarlos en problemas aplicados y 3. extender el estudio del comportamiento de funciones de agregación haciendo uso de variables aleatorias, han sido alcanzados.

No obstante, todavía hay algunos problemas abiertos. Dejando de lado la sección 6.5, en la que únicamente se han hecho algunos comentarios iniciales y un estudio en profundidad ha de hacerse en el futuro, las principales cuestiones son (1) ¿Existe alguna fórmula interpretable que pueda describir cualquier agregación de variables aleatorias en función de dos agregaciones de variables aleatorias, una de ellas condicionalmente determinada y la otra inducida aleatoriamente? (2) ¿Se cumple también la convergencia de los pesos óptimos acumulados cuando la distribución no es simétrica? (3) ¿Es posible dar la distribución exacta de la medida de un conjunto para capacidades aleatorias uniformes? (4) Si se considera el mismo problema que en la sección 4.5 pero con vectores aleatorios, ¿tiene solución el nuevo problema? (5) ¿Cuáles son las condiciones para tener el orden estocástico usual entre dos medidas de variabilidad muestrales cuando se considera un orden estocástico de dependencia entre los vectores aleatorios iniciales? y (6) ¿Es posible definir un orden estocástico de localización para conjuntos aleatorios que se preserve al aplicar uniones, intersecciones y productos cartesianos?

Dejando de lado estas preguntas aún sin respuesta, esta tesis pretende ser un primer paso en el uso de estructuras aleatorias en teoría de agregación cuando esta se aplica a análisis de datos. En este sentido, se recuerda una vez más, tal como fue explicado en la sección 1.3, la importancia del uso de herramientas probabilísticas y estadísticas en teoría de agregación. La probabilidad en agregación no es sólo necesaria, sino también indispensable.

List of Figures

List of Tables

References

[1] Gottfried Achenwall. Abriss der neuesten staatswissenschaft der heutigen vornehmsten europäischen reiche und republiken. Göttingen: Schmidt, 1749.

[2] Mohammad Ahsanullah and Ayman Alzaatreh. Parameter estimation for the log-logistic distribution based on order statistics. REVSTAT-Statistical Journal, 16(4):429–443, 2018.

[3] Ibrahim A Al-Kadit. Origins of cryptology: The arab contributions. Cryptologia, 16(2):97–126, 1992.

[4] Denis Allard, Alessandro Comunian, and Philippe Renard. Probability aggregation methods in geoscience. Mathematical Geosciences, 44:545–581, 2012.

[5] Héctor Allende and Carlos Valle. Ensemble methods for time series forecasting. In Claudio Moraga: A Passion for Multi-Valued Logic and Soft Computing, volume 349, pages 217–232. Springer, 2016.

[6] Hisham M Almongy, Ehab M Almetwally, Randa Alharbi, Dalia Alnagar, Eslam H Hafez, and Marwa M Mohie El-Din. The weibull generalized exponential distribution with censored sample: estimation and application on real data. Complexity, 2021(1):6653534, 2021.

[7] Douglas Altman, David Machin, Trevor Bryant, and Martin Gardner. Statistics with confidence: confidence intervals and statistical guidelines. John Wiley & Sons, New Jersey, USA, 2013.

[8] Omar Alzeley, Ehab M Almetwally, Ahmed M Gemeay, Huda M Alshanbari, Eslam H Hafez, and Mahmoud Hamed Abu-Moussa. Statistical inference under censored data for the new exponential-x fréchet distribution: Simulation and application to leukemia data. Computational Intelligence and Neuroscience, 2021(1):2167670, 2021.

[9] Takeshi Amemiya. Advanced econometrics. Harvard University Press, Cambridge, USA, 1985.

[10] Concezione Amitrano, Antonio Coniglio, and Francesco Di Liberto. Growth probability distribution in kinetic aggregation processes. Physical review letters, 57(8):1016, 1986.

[11] Donald WK Andrews. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. Econometrica, 70(1):119–162, 2002.

[12] Antonio Arriaza and Miguel A Sordo. Preservation of some stochastic orders by distortion functions with application to coherent systems with exchangeable components. Applied Stochastic Models in Business and Industry, 37(2):303–317, 2021.

[13] Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. Handbook of social choice and welfare. Elsevier, Amsterdam, The Netherlands, 2010.

[14] Alexandru V Asimit, Enkelejd Hashorva, and Dominik Kortschak. Aggregation of randomly weighted large risks. IMA Journal of Management Mathematics, 28(3):403–419, 2017.

[15] Jean-Pierre Aubin and Hélène Frankowska. Set-valued analysis. Springer Science & Business Media, Berlin, Germany, 2009.

[16] Narayanaswamy Balakrishnan, Yiying Zhang, and Peng Zhao. Ordering the largest claim amounts and ranges from two sets of heterogeneous portfolios. Scandinavian Actuarial Journal, 2018(1):23–41, 2018.

[17] Richard Baraniuk. Signals and systems, volume 8. Orange Grove Books, Johannesburg, South Africa, 2009.

[18] Ludwig Baringhaus and Rudolf Grübel. Nonparametric two-sample tests for increasing convex order. Bernoulli, 15(1):99–123, 2009.

[19] C Richard Barrett and Prasanta K Pattanaik. Aggregation of probability judgments. Econometrica, 55(5):1237–1241, 1987.

[20] Garry F Barrett and Stephen G Donald. Consistent tests for stochastic dominance. Econometrica, 71(1):71–104, 2003.

[21] Maciej Bartoszuk, Gleb Beliakov, Marek Gagolewski, and Simon James. Fitting aggregation functions to data: Part i-linearization and regularization. In Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 767–779. Springer, 2016.

[22] Maciej Bartoszuk, Gleb Beliakov, Marek Gagolewski, and Simon James. Fitting aggregation functions to data: Part ii-idempotization. In Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 780–789. Springer, 2016.

[23] Heinz Bauer. Measure and integration theory, volume 26. Walter de Gruyter, Boston, USA, 2001.

[24] Nicole Bäuerle. Inequalities for stochastic models via supermodular orderings. Stochastic Models, 13(1):181–201, 1997.

[25] Glen Baxter and Monroe D Donsker. On the distribution of the supremum functional for processes with stationary independent increments. Transactions of the American Mathematical Society, 85(1):73–87, 1957.

[26] Juan Baz, Gleb Beliakov, Irene Díaz, and Susana Montes. Taxonomy of aggregations of random variables. In Information Processing and Management of Uncertainty in Knowledge-Based Systems. In press. Springer, 2025.

[27] Juan Baz, Gleb Beliakov, Irene Díaz, and Susana Montes. Uniform fuzzy measures. Submitted, 2025.

[28] Juan Baz, Irene Díaz, Luis Garmendia, Daniel Gómez, Luis Magdalena, and Susana Montes. Computable aggregations of random variables. Information Sciences, 654:119842, 2024.

[29] Juan Baz, Irene Díaz, and Susana Montes. The choice of an appropriate stochastic order to aggregate random variables. In International Conference on Soft Methods in Probability and Statistics, pages 40–47. Springer, 2022.

[30] Juan Baz, Irene Díaz, and Susana Montes. Actualization probabilities of idempotent nullnorms and uninorms. In Intelligent Management of Data and Information in Decision Making: Proceedings of the 16th FLINS Conference on Computational Intelligence in Decision and Control & the 19th ISKE Conference on Intelligence Systems and Knowledge Engineering, pages 113–120. World Scientific, 2024.

[31] Juan Baz, Irene Díaz, and Susana Montes. Aggregation of random elements over bounded lattices. International Journal of Approximate Reasoning, 166:109122, 2024.

[32] Juan Baz, Irene Díaz, and Susana Montes. Minimizing the expected absolute difference with a fixed continuous copula. In International Conference on Soft Methods in Probability and Statistics, pages 18–26. Springer, 2024.

[33] Juan Baz, Irene Díaz, and Susana Montes. Idempotence and internality of aggregation of random variables. Submitted to EUSFLAT2025 conference, 2025.

[34] Juan Baz, Irene Díaz, and Susana Montes. Usual stochastic order for random sets. Submitted, 2025.

[35] Juan Baz, Mikel Ferrero-Jaurrieta, Irene Díaz, Susana Montes, Gleb Beliakov, and Humberto Bustince. Probabilistic study of induced ordered linear fusion operators for time series forecasting. Information Fusion, 103:102093, 2024.

[36] Juan Baz, Diego García-Zamora, Irene Díaz, Susana Montes, and Luis Martínez. Flexible-dimensional evr-owa as mean estimator for symmetric

distributions. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 11–24. Springer, 2022.

[37] Juan Baz, Diego García-Zamora, Irene Díaz, Susana Montes, and Luis Martínez. Flexible-dimensional l-statistic for mean estimation of symmetric distributions. Statistical Papers, 65:4001–4024, 2024.

[38] Juan Baz, Franco Pellerey, Irene Díaz, and Susana Montes. Stochastic ordering of variability measure estimators. Statistics, 58(1):26–43, 2024.

[39] Juan Baz, Franco Pellerey, Irene Díaz, and Susana Montes. Stochastically ordered aggregation operators. International Journal of Approximate Reasoning, 166:109110, 2024.

[40] Juan Baz, Franco Pellerey, Irene Díaz, and Susana Montes. On effects of dependence in variability estimation. Submitted, 2025.

[41] Juan Baz, Franco Pellerey, Irene Díaz, and Susana Montes. On the property of reduction of variabiltiy for mean functions. Submitted, 2025.

[42] G Beliakov. `Fmtools` package. `http://personal-sites.deakin.edu.au/$\sim$gleb/fmtools.html`, version 4, 2007-2021. [Online; accessed 18-November-2024].

[43] Gleb Beliakov. On random generation of supermodular capacities. IEEE Transactions on Fuzzy Systems, 30(1):293–296, 2020.

[44] Gleb Beliakov, Juan Baz, and Jian-Zhang Wu. Efficient random walks for generating random fuzzy measures in möbius representation in large universe. Computational and Applied Mathematics, 43(8):1–24, 2024.

[45] Gleb Beliakov, Humberto Bustince, and Tomasa Calvo. A practical guide to averaging functions, volume 329. Springer, New York, USA, 2016.

[46] Gleb Beliakov, Francisco Javier Cabrerizo, Enrique Herrera-Viedma, and Jian-Zhang Wu. Random generation of k-interactive capacities. Fuzzy Sets and Systems, 430:48–55, 2022.

[47] Gleb Beliakov, Tomasa Calvo, and Tim Wilkin. Three types of monotonicity of averaging functions. Knowledge-Based Systems, 72:114–122, 2014.

[48] Gleb Beliakov, Simon James, and J Wu. Discrete fuzzy measures. Springer, New York, USA, 2020.

[49] Gleb Beliakov, Ana Pradera, and Tomasa Calvo. Aggregation functions: A guide for practitioners, volume 221. Springer, New York, USA, 2007.

[50] Gleb Beliakov, Jana Špirková, and Tomasa Calvo. On the extended set of weights of the owa functions. International Journal of General Systems, 49(4):355–373, 2020.

[51] Gleb Beliakov and Jian-Zhang Wu. Random generation of capacities and its application in comprehensive decision aiding. Information Sciences, 577:424–435, 2021.

[52] Radim Belohlavek, Joseph W Dauben, and George J Klir. Fuzzy logic and mathematics: a historical perspective. Oxford University Press, Oxford, UK, 2017.

[53] Arie Beresteanu, Ilya Molchanov, and Francesca Molinari. Partial identification using random set theory. Journal of Econometrics, 166(1):17–32, 2012.

[54] Claude Berge. Topological spaces: including a treatment of multivalued functions, vector spaces and convexity. Oliver and Boyd, Edinburgh, Uk, 1963.

[55] Vance W Berger and YanYan Zhou. Kolmogorov–smirnov test: Overview. Wiley statsref: Statistics reference online, 2014.

[56] José R Berrendero, Antonio Cuevas, and Francisco Vjosázquez-grande. Testing multivariate uniformity: The distance-to-boundary method. Canadian Journal of Statistics, 34(4):693–707, 2006.

[57] James C Bezdek. Fuzziness vs. probability-again (!?). IEEE Transactions on Fuzzy systems, 2(1):1–3, 1994.

[58] Rajendra Bhatia. Positive definite matrices. Princeton university press, Princeton, USA, 2009.

[59] Peter J Bickel and Erich L Lehmann. Descriptive statistics for nonparametric models ii. location. In Selected Works of EL Lehmann, pages 473–497. Springer, 2011.

[60] Patrick Billingsley. Convergence of probability measures. John Wiley & Sons, New Jersey, USA, 2013.

[61] Yan Bin, Yu Hai-Bo, and Gao Zhen-Wei. A combination forecasting model based on iowa operator for dam safety monitoring. In 2013 Fifth International Conference on Measuring Technology and Mechatronics Automation, pages 5–8. IEEE, 2013.

[62] Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 4. Springer, New York, USA, 2006.

[63] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. Stochastic representation of fgm copulas using multivariate bernoulli random variables. Computational Statistics & Data Analysis, 173:107506, 2022.

[64] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. Risk aggregation with fgm copulas. Insurance: Mathematics and Economics, 111:102–120, 2023.

[65] Michał Boczek, Tomasz Józefiak, Marek Kaluszka, and Andrzej Okolewski. The choquet-like operator with respect to an admissible order as a tool for aggregating multivalued data. arXiv preprint arXiv:2404.07994, 2024.

[66] Vladimir Igorevich Bogachev and Maria Aparecida Soares Ruas. Measure theory, volume 1. Springer, New York, USA, 2007.

[67] Robert F Bordley and Ronald W Wolff. Note—on the aggregation of individual probability estimates. Management Science, 27(8):959–964, 1981.

[68] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. Time series analysis: forecasting and control. John Wiley & Sons, New Jersey, USA, 2015.

[69] Stephen Boyd and Lieven Vandenberghe. Convex optimization. Cambridge University Press, Cambridge, UK, 2004.

[70] James H Bray and Scott E Maxwell. Multivariate analysis of variance. Number 54 in Quantitative Applications in the Social Sciences. Sage, London, UK, 1985.

[71] Leo Breiman. Bagging predictors. Machine learning, 24:123–140, 1996.

[72] Zoltán Buczolich and Gábor J Székely. When is a weighted average of ordered sample elements a maximum likelihood estimator of the location parameter? Advances in Applied Mathematics, 10(4):439–456, 1989.

[73] Peter S Bullen. Handbook of means and their inequalities. Springer Science & Business Media, Berlin, Germany, 2013.

[74] Gregory L Bullock. A geometric interpretation of the riemann-stieltjes integral. The American Mathematical Monthly, 95(5):448–455, 1988.

[75] Humberto Bustince, Gleb Beliakov, Gracaliz Pereira Dimuro, Benjamin Bedregal, and Radko Mesiar. On the definition of penalty functions in data aggregation. Fuzzy Sets and Systems, 323:1–18, 2017.

[76] Humberto Bustince, Jose Antonio Sanz, Giancarlo Lucca, Graçaliz P Dimuro, Benjamin Bedregal, Radko Mesiar, Anna Kolesárová, and Gustavo Ochoa. Pre-aggregation functions: definition, properties and construction methods. In 2016 IEEE International Conference on Fuzzy Systems, pages 294–300. IEEE, 2016.

[77] Paul Cabilio and Joe Masaro. A simple test of symmetry about an unknown median. Canadian Journal of Statistics, 24(3):349–361, 1996.

[78] Tomasa Calvo and Gleb Beliakov. Aggregation functions based on penalties. Fuzzy Sets and Systems, 161(10):1420–1436, 2010.

[79] Luis M Candanedo, Véronique Feldheim, and Dominique Deramaix. Appliances Energy Prediction Dataset. `https://www.kaggle.com/datasets/loveall/appliances-energy-prediction`, 2017. [Online; accessed 01-May-2023].

[80] Luis M Candanedo, Véronique Feldheim, and Dominique Deramaix. Data driven prediction models of energy use of appliances in a low-energy house. Energy and buildings, 140:81–97, 2017.

[81] Marco Capaldo, Antonio Di Crescenzo, and Franco Pellerey. Generalized gini's mean difference through distortions and copulas, and related minimizing problems. Statistics & Probability Letters, 206:109981, 2024.

[82] Marco Capaldo and Jorge Navarro. New multivariate gini's indices. arXiv preprint arXiv:2401.01980, 2024.

[83] Ignacio Cascos, María Concepción López-Díaz, and Miguel López-Díaz. A stochastic order for interval valued random mappings and applications. Applied Mathematical Modelling, 107:429–440, 2022.

[84] Augustin Louis Baron Cauchy. Cours d'analyse de l'École Royale Polytechnique. Imprimerie Royale, Paris, France, 1821.

[85] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. The Journal of Machine Learning Research, 11:2079–2107, 2010.

[86] Vitor Cerqueira, Luis Torgo, Mariana Oliveira, and Bernhard Pfahringer. Dynamic and heterogeneous ensembles for time series forecasting. In 2017 IEEE international conference on data science and advanced analytics, pages 242–251. IEEE, 2017.

[87] Peter Cheeseman. Probabilistic versus fuzzy reasoning. In Machine Intelligence and Pattern Recognition, volume 4, pages 85–102. Elsevier, 1986.

[88] Kyunghyun Cho. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.

[89] Danbi Chung, Oliver Linton, and Yoon-Jae Whang. Consistent testing for an implication of supermodular dominance. Cambridge Working Papers in Economics, (2143), 2021.

[90] Kai Lai Chung. Probability and doob. The American mathematical monthly, 105(1):28–35, 1998.

[91] Robert T Clemen. Calibration and the aggregation of probabilities. Management Science, 32(3):312–314, 1986.

[92] Donald L Cohn. Measurable choice of limit points and the existence of separable and measurable processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 22(2):161–165, 1972.

[93] Donald L Cohn. Measure theory, volume 1. Springer, New York, USA, 2013.

[94] Ana Colubi. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy-and real-valued data. Fuzzy Sets and Systems, 160(3):344–356, 2009.

[95] Elías F Combarro, Julen Hurtado de Saracho, and Irene Díaz. Minimals plus: an improved algorithm for the random generation of linear extensions of partially ordered sets. Information Sciences, 501:50–67, 2019.

[96] Elías F Combarro, Irene Díaz, and Pedro Miranda. On random generation of fuzzy measures. Fuzzy Sets and Systems, 228:64–77, 2013.

[97] Areski Cousin and Jean-Paul Laurent. Comparison results for exchangeable credit risk portfolios. Insurance: Mathematics and Economics, 42(3):1118–1127, 2008.

[98] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1):21–27, 1967.

[99] Harald Cramér. A contribution to the theory of statistical estimation. Scandinavian Actuarial Journal, 1946(1):85–94, 1946.

[100] Ernest Czogała and Józef Drewniak. Associative monotonic operations in fuzzy set theory. Fuzzy Sets and Systems, 12(3):249–269, 1984.

[101] Karma Dajani and Sjoerd Dirksin. A simple introduction to ergodic theory. University of Utrecht, Lecture notes in Ergodic Theory, 2008.

[102] Brian A Davey and Hilary A Priestley. Introduction to Lattices and Order. Cambridge University Press, Cambridge, UK, 2002.

[103] Herbert A David. Miscellanea: Gini's mean difference rediscovered. Biometrika, 55(3):573–575, 1968.

[104] Herbert A David and Haikady N Nagaraja. Order statistics. John Wiley & Sons, New Jersey, USA, 2004.

[105] Enrique de Amo, M Díaz Carrillo, and Juan Fernández-Sánchez. Characterization of all copulas associated with non-continuous random variables. Fuzzy Sets and Systems, 191:103–112, 2012.

[106] Bernard De Baets, Hans De Meyer, and Radko Mesiar. Lipschitz continuity of copulas wrt lp-norms. Nonlinear Analysis: Theory, Methods & Applications, 72(9-10):3722–3731, 2010.

[107] Laécio Carvalho De Barros, Rodney Carlos Bassanezi, and Weldon Alexander Lodwick. First Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics. Springer, New York, USA, 2016.

[108] Miguel De Carvalho. Mean, what do you mean? The American Statistician, 70(3):270–274, 2016.

[109] Lokenath Debnath and Kanadpriya Basu. A short history of probability theory and its applications. International Journal of Mathematical Education in Science and Technology, 46(1):13–39, 2015.

[110] Ana del Amo, Javier Montero, and Elisenda Molina. Representation of consistent recursive rules. European Journal of Operational Research, 130(1):29–53, 2001.

[111] Luc Devroye. Non-uniform random variate generation. Handbooks in operations research and management science, 13:83–121, 2006.

[112] Camila Dias, Jessica Bueno, Eduardo Borges, Giancarlo Lucca, Helida Santos, Graçaliz Dimuro, Humberto Bustince, Paulo Drews, Silvia Botelho, and Eduardo Palmeira. Simulating the behaviour of choquet-like (pre) aggregation functions for image resizing in the pooling layer of deep learning networks. In Fuzzy Techniques: Theory and Applications: Proceedings of the 2019 Joint World Congress of the International Fuzzy Systems Association and the Annual Conference of the North American Fuzzy Information Processing Society, pages 224–236. Springer, 2019.

[113] Susana Diaz-Vazquez, Emilio Torres-Manzanera, Noelia Rico, Radko Mesiar, Iosu Rodriguez-Martinez, Julio Lafuente, Irene Diaz, Susana Montes, and Humberto Bustince. A new family of aggregation functions for intervals. Computational and Applied Mathematics, 43(1):17, 2024.

[114] Graçaliz Pereira Dimuro, Javier Fernández, Benjamin Bedregal, Radko Mesiar, José Antonio Sanz, Giancarlo Lucca, and Humberto Bustince. The state-of-art of the generalizations of the choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions. Information Fusion, 57:27–43, 2020.

[115] Peng Ding. Three occurrences of the hyperbolic-secant distribution. The American Statistician, 68(1):32–35, 2014.

[116] Didier Dubois, Weiru Liu, Jianbing Ma, and Henri Prade. The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. Information Fusion, 32:12–39, 2016.

[117] Alexander Dukhovny. Lattice polynomials of random variables. Statistics & Probability Letters, 77(10):989–994, 2007.

[118] Fabrizio Durante and Carlo Sempi. Principles of copula theory, volume 474. CRC press, Boca Raton, USA, 2016.

[119] Rick Durrett. Probability: theory and examples, volume 49. Cambridge University Press, Cambridge, UK, 2019.

[120] Alex Dytso, Ronit Bustin, H Vincent Poor, and Shlomo Shamai. Analytical properties of generalized gaussian distributions. Journal of Statistical Distributions and Applications, 5:1–40, 2018.

[121] Alex Dytso, Martina Cardone, Mishfad S Veedu, and H Vincent Poor. On estimation under noisy order statistics. In 2019 IEEE International Symposium on Information Theory, pages 36–40. IEEE, 2019.

[122] Ulle Endriss and Umberto Grandi. Graph aggregation. In Companion Proceedings of the The Web Conference 2018, pages 447–450, 2018.

[123] Paul Erdős and Alfréd Renyi. On random graphs. Publicationes Mathematicae, 6:290–297, 1959.

[124] Bard Ermentrout and Ajay Mahajan. Simulating, analyzing, and animating dynamical systems: a guide to xppaut for researchers and students. Appl. Mech. Rev., 56(4):B53–B53, 2003.

[125] Yanqin Fan and Marc Henry. Vector copulas. Journal of Econometrics, 234(1):128–150, 2023.

[126] Alireza Faraz and Arnold F Shapiro. An application of fuzzy random variables to control charts. Fuzzy Sets and Systems, 161(20):2684–2694, 2010.

[127] Ignacio Cascos Fernández and Ilya Molchanov. A stochastic order for random vectors and random sets based on the aumann expectation. Statistics & Probability Letters, 63(3):295–305, 2003.

[128] Peter C Fishburn and Ariel Rubinstein. Aggregation of equivalence relations. Journal of classification, 3:61–65, 1986.

[129] Roberto Fontana and Patrizia Semeraro. Exchangeable bernoulli distributions: High dimensional simulation, estimation, and testing. Journal of Statistical Planning and Inference, 225:52–70, 2023.

[130] Maurice Fréchet. Sur l'extension de certaines évaluations statistiques au cas de petits échantillons. Revue de l'Institut International de Statistique, 11(3/4):182–205, 1943.

[131] Jerome H Friedman. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5):1189–1232, 2001.

[132] Jerome H Friedman. Stochastic gradient boosting. Computational statistics & data analysis, 38(4):367–378, 2002.

[133] Esther Frostig. Comparison of portfolios which depend on multivariate bernoulli random variables with fixed marginals. Insurance: Mathematics and Economics, 29(3):319–331, 2001.

[134] Javier Fumanal-Idocin, Zdenko Takáč, Javier Fernández, José Antonio Sanz, Harkaitz Goyena, Ching-Teng Lin, Yu-Kai Wang, and Humberto Bustince. Interval-valued aggregation functions based on moderate deviations applied to motor-imagery-based brain–computer interface. IEEE Transactions on Fuzzy Systems, 30(7):2706–2720, 2021.

[135] Marek Gagolewski. Penalty-based aggregation of multidimensional data. Fuzzy Sets and Systems, 325:4–20, 2017.

[136] Marek Gagolewski, Raúl Pérez-Fernández, and Bernard De Baets. An inherent difficulty in the aggregation of multidimensional data. IEEE Transactions on Fuzzy Systems, 28(3):602–606, 2019.

[137] Alex Gaina. Details on biography of jerzy neyman. Moldavian Journal of the physical Sciences, 2(3-4):393–395, 2003.

[138] Wei Gao, Teng Zhang, Bin-Bin Yang, and Zhi-Hua Zhou. On the noise estimation statistics. Artificial Intelligence, 293:103451, 2021.

[139] Diego García-Zamora, Álvaro Labella, Rosa M Rodríguez, and Luis Martínez. Nonlinear preferences in group decision-making. extreme values amplifications and extreme values reductions. International Journal of Intelligent Systems, 36(11):6581–6612, 2021.

[140] Diego García-Zamora, Álvaro Labella, Rosa M Rodríguez, and Luis Martínez. Symmetric weights for owa operators prioritizing intermediate values. the evr-owa operator. Information Sciences, 584:583–602, 2022.

[141] David JH Garling. Analysis on Polish spaces and an introduction to optimal transportation, volume 89. Cambridge University Press, Cambridge, UK, 2018.

[142] Gene H Golub and Charles F Van LOAN. Cf matrix computations, johns hopkins studies in the mathematical sciences. The Johns Hopkins University Press, Baltimore, MD, USA, 90:93–94, 1996.

[143] Daniel Gómez and Javier Montero. A discussion on aggregation operators. Kybernetika, 40(1):107–120, 2004.

[144] John Goutsias, Ronald PS Mahler, and Hung T Nguyen. Random sets: theory and applications, volume 97. Springer Science & Business Media, Berlin, Germany, 2012.

[145] Michel Grabisch. The application of fuzzy integrals in multicriteria decision making. European journal of operational research, 89(3):445–456, 1996.

[146] Michel Grabisch. Set functions, games and capacities in decision making, volume 46. Springer, New York, USA, 2016.

[147] Michel Grabisch and Christophe Labreuche. A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. Annals of Operations Research, 175:247–286, 2010.

[148] Michel Grabisch, Christophe Labreuche, and Peiqi Sun. An approximation algorithm for random generation of capacities. Order, 41:281–306, 2023.

[149] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. Aggregation functions, volume 127. Cambridge University Press, Cambridge, UK, 2009.

[150] Michel Grabisch and Eric Raufaste. An empirical study of statistical properties of the choquet and sugeno integrals. IEEE Transactions on Fuzzy Systems, 16(4):839–850, 2008.

[151] Michel Grabisch and Marc Roubens. Application of the choquet integral in multicriteria decision making. In Fuzzy Measures and Integrals-Theory and Applications, pages 348–374. Springer, New York, USA, 2000.

[152] Benjamin T Graham and Geoffrey R Grimmett. Influence and sharp-threshold theorems for monotonic measures. The Annals of Probability, 34(5):1726 – 1745, 2006.

[153] Geoffrey Grimmett and David Stirzaker. Probability and random processes. Oxford University Press, Oxford, UK, 2020.

[154] M Guojian, Bo Zhong, and Zhang Xiangbo. Combination forecasting model of equipment spare parts demand based on iowa operator. Ordnance Industry Automation, 32(1):8–11, 2013.

[155] William W Hager. Updating the inverse of a matrix. SIAM review, 31(2):221–239, 1989.

[156] James D Hamilton. Time series analysis. Princeton university press, Princeton, USA, 2020.

[157] Amal S Hassan and Marwa Abd-Allah. Exponentiated weibull-lomax distribution: properties and estimation. Journal of data science, 16(2):277–298, 2018.

[158] Kaijian He, Qian Yang, Lei Ji, Jingcheng Pan, and Yingchao Zou. Financial time series forecasting with the deep learning ensemble model. Mathematics, 11(4):1054, 2023.

[159] Jeff Henrikson. Completeness and total boundedness of the hausdorff metric. MIT Undergraduate Journal of Mathematics, 1(69-80):10, 1999.

[160] Tin Kam Ho. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition, volume 1, pages 278–282. IEEE, 1995.

[161] Sepp Hochreiter, Jürgen Schmidhuber, and Corso Elvezia. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.

[162] Richard Hoffman and Anil K Jain. A test of randomness based on the minimal spanning tree. Pattern Recognition Letters, 1(3):175–180, 1983.

[163] Stephen C Hora, Benjamin R Fransen, Natasha Hawkins, and Irving Susel. Median aggregation of distribution functions. Decision Analysis, 10(4):279–291, 2013.

[164] Jonathan RM Hosking. L-estimation. Handbook of Statistics, 17:215–235, 1998.

[165] Hsi-Mei Hsu and Chen-Tung Chen. Aggregation of fuzzy opinions under group decision making. Fuzzy Sets and Systems, 79(3):279–285, 1996.

[166] Joachim Inkmann. Asymptotic properties of gmm estimators. In Conditional Moment Estimation of Nonlinear Equation Systems: With an Application to an Oligopoly Model of Cooperative R&D, pages 20–27. Springer, New York, USA, 2001.

[167] Muhammad Aminul Islam, Derek T Anderson, Fred Petry, and Paul Elmore. An efficient evolutionary algorithm to optimize the choquet integral. International Journal of Intelligent Systems, 34(3):366–385, 2019.

[168] Kurt Jacobs. Stochastic processes for physicists: understanding noisy systems. Cambridge University Press, Cambridge, UK, 2010.

[169] Anil K Jain, Xiaowei Xu, Tin Kam Ho, and Fan Xiao. Uniformity testing using minimal spanning tree. In 2002 International Conference on Pattern Recognition, volume 4, pages 281–284. IEEE, 2002.

[170] Simon James. An introduction to data analysis using aggregation functions in R. Springer, New York, USA, 2016.

[171] Johan Ludwig William Valdemar Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta mathematica, 30(1):175–193, 1906.

[172] Vilde Jensen, Filippo Maria Bianchi, and Stian Normann Anfinsen. Ensemble conformalized quantile regression for probabilistic time series forecasting. IEEE Transactions on Neural Networks and Learning Systems, 2022.

[173] Jiteng Jia, Tomohiro Sogabe, and Moawwad El-Mikkawy. Inversion of k-tridiagonal matrices with toeplitz structure. Computers & Mathematics with Applications, 65(1):116–125, 2013.

[174] Guorui Jiang and Ying Liu. Research on collaborative forecasting model based on cpfr. In Software Engineering and Knowledge Engineering: Theory and Practice: Volume 1, pages 523–529. Springer, 2012.

[175] Vesselin P Jilkov and X Rong Li. Online bayesian estimation of transition probabilities for markovian jump systems. IEEE Transactions on signal processing, 52(6):1620–1630, 2004.

[176] Norman L Johnson, Adrienne W Kemp, and Samuel Kotz. Univariate discrete distributions, volume 444. John Wiley & Sons, New Jersey, USA, 2005.

[177] Norman L Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. Continuous univariate distributions, volume 2. John Wiley & Sons, New Jersey, USA, 1995.

[178] Oscar Jorda. Random-time aggregation in partial adjustment models. Journal of Business & Economic Statistics, 17(3):382–395, 1999.

[179] Anosh Joseph. Markov chain monte carlo methods in quantum field theories: A modern primer. Springer Nature, London, UK, 2020.

[180] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. Likert scale: Explored and explained. British Journal of Applied Science & Technology, 7(4):396–403, 2015.

[181] Hans-Michael Kaltenbach. A concise guide to statistics. Springer Science & Business Media, Berlin, Germany, 2011.

[182] Funda Karacal and Radko Mesiar. Aggregation functions on bounded lattices. International Journal of General Systems, 46(1):37–51, 2017.

[183] Paweł Karczmarek, Michał Dolecki, Paweł Powroznik, Zbigniew A Łagodowski, Adam Gregosiewicz, Łukasz Gałka, Witold Pedrycz, Dariusz Czerwinski, and Kamil Jonak. Quadrature-inspired generalized choquet integral in an application to classification problems. IEEE Access, 2023.

[184] H Jerome Keisler. An infinitesimal approach to stochastic analysis, volume 297. American Mathematical Soc., Providence, USA, 1984.

[185] H Jerome Keisler. Hyperfinite models of adapted probability logic. Annals of Pure and Applied Logic, 31:71–86, 1986.

[186] H Jerome Keisler. Rich and saturated adapted spaces. Advances in Mathematics, 128(2):242–288, 1997.

[187] H Jerome Keisler and Yeneng Sun. Why saturated probability spaces are necessary. Advances in Mathematics, 221(5):1584–1607, 2009.

[188] John Frank Charles Kingman. Poisson processes, volume 3. Clarendon Press, Oxford, UK, 1992.

[189] John P Klein and Melvin L Moeschberger. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media, Berlin, Germany, 2006.

[190] Erich Peter Klement, Radko Mesiar, and Endre Pap. Triangular norms, volume 8. Springer Science & Business Media, Berlin, Germany, 2013.

[191] Achim Klenke. Probability theory: a comprehensive course. Springer Science & Business Media, Berlin, Germany, 2013.

[192] George J Klir and Tina A Folger. Fuzzy sets, uncertainty, and information. Prentice-Hall, Inc., Denver, USA, 1987.

[193] Ivan Kojadinovic and Jean-Luc Marichal. On the moments and distribution of discrete choquet integrals from continuous distributions. Journal of computational and applied mathematics, 230(1):83–94, 2009.

[194] Andrey Kolmogoroff. Grundbegriffe der wahrscheinlichkeitsrechnung. Springer, New York, USA, 1933.

[195] Andréi Nikolaevich Kolmogorov and Albert T Bharucha-Reid. Foundations of the theory of probability: Second English Edition. Courier Dover Publications, Paris, France, 2018.

[196] Andreĭ Nikolaevich Kolmogorov and Guido Castelnuovo. Sur la notion de la moyenne. G. Bardi, tip. della R. Accad. dei Lincei, Rome, Italy, 1930.

[197] Magda Komornikova and Radko Mesiar. Aggregation functions on bounded partially ordered sets and their classification. Fuzzy Sets and Systems, 175(1):48–56, 2011.

[198] Thaleia Kontoroupi and Andrew W Smyth. Online noise identification for joint state and parameter estimation of nonlinear systems. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 2(3):B4015006, 2016.

[199] Milos Kopa and Barbora Petrová. Strong and weak multivariate first-order stochastic dominance. Available at SSRN 3144058, 2018.

[200] Bart Kosko. Fuzziness vs. probability. International Journal of General System, 17(2-3):211–240, 1990.

[201] Samuel Kotz, Narayanaswamy Balakrishnan, and Norman L Johnson. Continuous multivariate distributions, Volume 1: Models and applications. John Wiley & Sons, New Jersey, USA, 2019.

[202] William Kruskal. The significance of fisher: a review of ra fisher: the life of a scientist. Journal of the American Statistical Association, 75(372):1019–1030, 1980.

[203] Devendra Kumar, Maneesh Kumar, and JP Singh Joorel. Estimation with modified power function distribution based on order statistics with application to evaporation data. Annals of Data Science, 9(4):723–748, 2022.

[204] Soumendra N Lahiri. Theoretical comparisons of block bootstrap methods. Annals of Statistics, 27:386–404, 1999.

[205] Pierre Simon Laplace. Théorie analytique des probabilités. Courcier Dover Publications, Paris, France, 1820.

[206] Daniel Leite and Igor Škrjanc. Ensemble of evolving optimal granular experts, owa aggregation, and time series prediction. Information sciences, 504:95–112, 2019.

[207] Bo Li, Junqi Ding, Zhengqing Yin, Kaiyu Li, Xue Zhao, and Lingxian Zhang. Optimized neural network combined model based on the induced ordered weighted averaging operator for vegetable price forecasting. Expert Systems with Applications, 168:114232, 2021.

[208] Fen Li, Jialin Qian, Quanquan Yan, Xingwu Yang, Jinbin Zhao, Keqing Qu, and Qijun Song. A combination forecasting model based on iowa operator for pv generation. In International Conference on Renewable Power Generation, pages 1–4. IET, 2015.

[209] Guohui Li, Wanni Chang, and Hong Yang. A novel combined prediction model for monthly mean precipitation with error correction strategy. IEEE Access, 8:141432–141445, 2020.

[210] Di Liang, Long Fei Ma, Ya Feng Hu, and Shuang Wu. Combination forecasting method based on iowa operator and application. Advanced Materials Research, 945:2515–2518, 2014.

[211] Thomas Milton Liggett. Continuous time Markov processes: an introduction. American Mathematical Society, Providence, USA, 2010.

[212] Dennis V Lindley. Scoring rules and the inevitability of probability. International Statistical Review/Revue Internationale de Statistique, 50:1–11, 1982.

[213] James K Lindsey. Statistical analysis of stochastic processes in time, volume 14. Cambridge University Press, Cambridge, UK, 2004.

[214] Roderick JA Little and Donald B Rubin. Statistical analysis with missing data, volume 793. John Wiley & Sons, Hoboken, New Jersey, 2019.

[215] Gang Liu, Fuyuan Xiao, Chin-Teng Lin, and Zehong Cao. A fuzzy interval time-series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced-ordered weighted averaging aggregation operation. IEEE Transactions on Fuzzy Systems, 28(11):2677–2690, 2020.

[216] Xiao Liu, Amol Ghorpade, YL Tu, and Wen-Jun Zhang. A novel approach to probability distribution aggregation. Information Sciences, 188:269–275, 2012.

[217] Xinsheng Liu and Jinde Wang. Testing for increasing convex order in several populations. Annals of the Institute of Statistical Mathematics, 55:121–136, 2003.

[218] Emlyn H Lloyd. Least-squares estimation of location and scale parameters using order statistics. Biometrika, 39(1/2):88–95, 1952.

[219] Zadeh Lotfi A. Fuzzy sets. Information and control, 8:35–40, 1965.

[220] Giancarlo Lucca, Jose Sanz, Graçaliz P Dimuro, Benjamín Bedregal, and Humberto Bustince. Analyzing the behavior of aggregation and pre-aggregation functions in fuzzy rule-based classification systems with data complexity measures. In Advances in Fuzzy Logic and Technology 2017:

Proceedings of: EUSFLAT-2017–The 10th Conference of the European Society for Fuzzy Logic and Technology, pages 443–455. Springer, 2018.

[221] Giancarlo Lucca, José Antonio Sanz, Humberto Bustince, Graçaliz P Dimuro, Veronique Gomes, Rui Cláudio Constantino Madureira, and Pedro Melo-Pinto. Applying aggregation and pre-aggregation functions in the classification of grape berries. In 2018 IEEE International Conference on Fuzzy Systems, pages 1–6. IEEE, 2018.

[222] Eugene Lukacs. Stochastic convergence. Academic Press, London, UK, 2014.

[223] Aidan Lyon, Bonnie C Wintle, and Mark Burgman. Collective wisdom: Methods of confidence interval aggregation. Journal of Business Research, 68(8):1759–1767, 2015.

[224] Kanti V Mardia, John T Kent, and Charles C Taylor. Multivariate analysis, volume 88. John Wiley & Sons, New Jersey, USA, 2024.

[225] James E Marengo, David L Farnsworth, and Lucas Stefanic. A geometric derivation of the irwin-hall distribution. International Journal of Mathematics and Mathematical Sciences, 2017(1):3571419, 2017.

[226] Jean-Luc Marichal. Tolerant or intolerant character of interacting criteria in aggregation by the choquet integral. European Journal of Operational Research, 155(3):771–791, 2004.

[227] Jean-Luc Marichal. Cumulative distribution functions and moments of lattice polynomials. Statistics & Probability Letters, 76(12):1273–1279, 2006.

[228] Jean-Luc Marichal. Weighted lattice polynomials of independent random variables. Discrete Applied Mathematics, 156(5):685–694, 2008.

[229] Jean-Luc Marichal. Aggregation functions for decision making. arXiv preprint arXiv:0901.4232, 2009.

[230] Jean-Luc Marichal and Ivan Kojadinovic. Distribution functions of linear combinations of lattice polynomials from the uniform distribution. Statistics & Probability Letters, 78(8):985–991, 2008.

[231] Albert W Marshall and Ingram Olkin. Families of multivariate distributions. Journal of the American statistical association, 83(403):834–841, 1988.

[232] Daniel McFadden. Testing for stochastic dominance. In Studies in the economics of uncertainty: In honor of Josef Hadar, pages 113–134. Springer, 1989.

[233] Geoffrey J McLachlan, Sharon X Lee, and Suren I Rathnayake. Finite mixture models. Annual review of statistics and its application, 6(1):355–378, 2019.

[234] R Mesiar and A Kolesarova. On the fuzzy set theory and aggregation functions: History and some recent advances. Iranian Journal of Fuzzy Systems, 15(7):1–12, 2018.

[235] Thadeu Milfont, Ivan Mezzomo, Benjamin Bedregal, Edmundo Mansilla, and Humberto Bustince. Aggregation functions on n-dimensional ordered vectors equipped with an admissible order and an application in multi-criteria group decision-making. International Journal of Approximate Reasoning, 137:34–50, 2021.

[236] Enrique Miranda and Hung T Nguyen. Random sets and imprecise probabilities. International Journal of Approximate Reasoning, 46(1):1–2, 2007.

[237] Pedro Miranda and Elías F Combarro. On the structure of some families of fuzzy measures. IEEE Transactions on Fuzzy Systems, 15(6):1068–1081, 2007.

[238] Wan Rosanisah Wan Mohd and Lazim Abdullah. Aggregation methods in group decision making: A decade survey. Informatica, 41(1), 2017.

[239] Ilya Molchanov and Ilya S Molchanov. Theory of random sets, volume 19. Springer, New York, USA, 2005.

[240] Ilya Molchanov and Francesca Molinari. Applications of random set theory in econometrics. Annual Review of Economics, 6(1):229–251, 2014.

[241] Javier Montero, Ramón González-del Campo, Luis Garmendia, Daniel Gómez, and J Tinguaro Rodríguez. Computable aggregations. Information Sciences, 460:439–449, 2018.

[242] Ignacio Montes, Enrique Miranda, and Susana Montes. Stochastic dominance with imprecise information. Computational Statistics & Data Analysis, 71:868–886, 2014.

[243] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. Introduction to time series analysis and forecasting. John Wiley & Sons, New Jersey, USA, 2015.

[244] Julio Mulero, Franco Pellerey, and Rosario Rodríguez-Griñolo. Stochastic comparisons for time transformed exponential models. Insurance: Mathematics and Economics, 46(2):328–333, 2010.

[245] Alfred Müller. Stochastic orders generated by integrals: a unified study. Advances in Applied probability, 29(2):414–428, 1997.

[246] Alfred Müller. On the waiting times in queues with dependency between interarrival and service times. Operations Research Letters, 26(1):43–47, 2000.

[247] Alfred Müller. Stochastic ordering of multivariate normal distributions. Annals of the Institute of Statistical Mathematics, 53:567–575, 2001.

[248] James R Munkres. Topology. Prentice Hall, Hoboken, USA, 2000.

[249] Muhammad Muzammal, Romana Talat, Ali Hassan Sodhro, and Sandeep Pirbhulal. A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. Information Fusion, 53:155–164, 2020.

[250] Naveen Narisetty and Roger Koenker. Censored quantile regression survival models with a cure proportion. Journal of Econometrics, 226(1):192–203, 2022.

[251] Robert F Nau. The aggregation of imprecise probabilities. Journal of Statistical Planning and Inference, 105(1):265–282, 2002.

[252] Jorge Navarro and Yolanda del Águila. Stochastic comparisons of distorted distributions, coherent systems and mixtures with ordered components. Metrika, 80:627–648, 2017.

[253] Roger B Nelsen. An introduction to copulas. Springer, New York, USA, 2006.

[254] Hung T Nguyen. An introduction to random sets. CRC press, Boca Raton, USA, 2006.

[255] James R Norris. Markov chains. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK, 1998.

[256] Ilkka Norros. A compensator representation of multivariate life length distributions, with applications. Scandinavian Journal of Statistics, 13(2):99–112, 1986.

[257] Paulo Eduardo Oliveira. Asymptotics for associated random variables. Springer Science & Business Media, Berlin, Germany, 2012.

[258] Patricia Ortega-Jiménez, Franco Pellerey, Miguel A Sordo, and Alfonso Suárez-Llorens. A minimizing problem of distances between random variables with proportional reversed hazard rate functions. In International Conference on Soft Methods in Probability and Statistics, pages 311–318. Springer, 2022.

[259] Patricia Ortega-Jiménez, Miguel A Sordo, and Alfonso Suárez-Llorens. Stochastic comparisons of some distances between random variables. Mathematics, 9(9):981, 2021.

[260] Walenty Ostasiewicz. The emergence of statistical science. Śląski Przegląd Statystyczny, 18(12):75–82, 2014.

[261] Athanasios Papoulis and S Unnikrishna Pillai. Probability, random variables, and stochastic processes, Fourth Edition. McGraw-Hill Europe, New York, USA, 2002.

[262] Kalyanapuram Rangachari Parthasarathy. Probability measures on metric spaces, volume 352. American Mathematical Soc., Providence, USA, 2005.

[263] Daniel Paternain, Laura De Miguel, Gustavo Ochoa, Inmaculada Lizasoain, Radko Mesiar, and Humberto Bustince. The interval-valued choquet integral based on admissible permutations. IEEE Transactions on Fuzzy Systems, 27(8):1638–1647, 2018.

[264] Daniel Paternain, LeSheng Jin, Radko Mesiar, Lucia Vavríková, and Humberto Bustince. Owa operators based on admissible permutations. In 2019 IEEE International Conference on Fuzzy Systems, pages 1–5. IEEE, 2019.

[265] Raúl Pérez-Fernández. Multivariate OWA functions. Fuzzy Sets and Systems, 466:108419, 2023.

[266] Raúl Pérez-Fernández and Bernard De Baets. Aggregation theory revisited. IEEE Transactions on Fuzzy Systems, 29(4):797–804, 2020.

[267] Raúl Pérez-Fernández, Gustavo Ochoa, Susana Montes, Irene Díaz, Javier Fernández, Daniel Paternain, and Humberto Bustince. Axiomatization and construction of orness measures for aggregation functions. International Journal of Intelligent Systems, 36(5):2208–2228, 2021.

[268] Karl E Petersen. Ergodic theory, volume 2. Cambridge University Press, Cambridge, UK, 1989.

[269] Yann Petot, Pierre Vallois, and Alexandre Voisin. Choquet integral with stochastic entries. Fuzzy Sets and Systems, 412:80–94, 2021.

[270] Adam Petrie and Thomas R Willemain. An empirical study of tests for uniformity in multidimensional data. Computational Statistics & Data Analysis, 64:253–268, 2013.

[271] Francesco Piccialli, Fabio Giampaolo, Edoardo Prezioso, David Camacho, and Giovanni Acampora. Artificial intelligence and healthcare: Forecasting of medical bookings through multi-source time-series fusion. Information Fusion, 74:1–16, 2021.

[272] Madan L Puri and Dan A Ralescu. Fuzzy random variables. Journal of Mathematical Analysis and Applications, 114:409–422, 1986.

[273] Xueheng Qiu, Ponnuthurai Nagaratnam Suganthan, and Gehan AJ Amaratunga. Fusion of multiple indicators with ensemble incremental learning techniques for stock price forecasting. Journal of Banking and Financial Technology, 3:33–42, 2019.

[274] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[275] Dan Ralescu. Cardinality, quantifiers, and the aggregation of fuzzy criteria. Fuzzy Sets and Systems, 69(3):355–365, 1995.

[276] C Radhakrishna Rao. Information and accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society, 37(3):81–91, 1945.

[277] R Tyrrell Rockafellar. Lagrange multipliers and optimality. SIAM review, 35(2):183–238, 1993.

[278] Vijay K Rohatgi and AK Md Ehsanes Saleh. An introduction to probability and statistics. John Wiley & Sons, New Jersey, USA, 2015.

[279] Julián Ros, Raúl Pérez-Fernández, and Ignacio Montes. Comparing bivariate random vectors by means of statistical preference. In International Conference on Soft Methods in Probability and Statistics, pages 396–404. Springer, 2024.

[280] David A Ross. Loeb measure and probability. In Nonstandard Analysis: Theory and Applications, pages 91–120. Springer Netherlands, Dordrecht, 1997.

[281] Timothy J Ross, Jane M Booker, and W Jerry Parkinson. Fuzzy logic and probability applications: bridging the gap. SIAM, Philadelphia, USA, 2002.

[282] Havard Rue and Leonhard Held. Gaussian Markov random fields: theory and applications. CRC press, Boca Raton, USA, 2005.

[283] Dymitr Ruta and Bogdan Gabrys. Neural network ensembles for time series prediction. In 2007 International Joint Conference on Neural Networks, pages 1204–1209. IEEE, 2007.

[284] Omer Sagi and Lior Rokach. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4):1249, 2018.

[285] Wolfgang Sander. Associative aggregation operators. In Aggregation Operators: New Trends and Applications, pages 124–158. Physica-Verlag HD, Heidelberg, Germany, 2002.

[286] Ahmed E Sarhan. Estimation of the mean and standard deviation by order statistics. The Annals of Mathematical Statistics, 25(2):317–328, 1954.

[287] Ahmed E Sarhan. Estimation of the mean and standard deviation by order statistics. part II. The Annals of Mathematical Statistics, 26(3):505–511, 1955.

[288] Ahmed E Sarhan. Estimation of the mean and standard deviation by order statistics. part III. The Annals of Mathematical Statistics, 26(4):576–592, 1955.

[289] Dirk Schlimm. Axioms in mathematical practice. Philosophia Mathematica, 21(1):37–92, 2013.

[290] Hans Schneider and George Phillip Barker. Matrices and linear algebra. Courier Dover Publications, Paris, France, 1989.

[291] Paul J Schweitzer. A survey of aggregation-disaggregation in large markov chains. In Numerical solution of Markov chains, pages 63–88. CRC press, Boca Raton, USA, 2021.

[292] Berthold Schweizer and Abe Sklar. Probabilistic metric spaces. Courier Dover Publications, Paris, France, 2011.

[293] Mikel Sesma-Sara, Humberto Bustince, and Radko Mesiar. Directional monotonicity of multidimensional fusion functions with respect to admissible orders. Fuzzy Sets and Systems, 467:108498, 2023.

[294] Glenn Shafer. The significance of jacob bernoulli's ars conjectandi for the philosophy of probability today. Journal of Econometrics, 75(1):15–32, 1996.

[295] Moshe Shaked and J George Shanthikumar. Stochastic orders. Springer, New York, USA, 2007.

[296] OB Sheynin. Carl friedrich gauss: b. 30 april 1777 d. 23 february 1855. In Statisticians of the Centuries, pages 119–122. Springer, New York, USA, 2001.

[297] Galen R Shorack and Jon A Wellner. Empirical processes with applications to statistics. SIAM, Philadelphia, USA, 2009.

[298] Abe Sklar. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris, 8:229–231, 1959.

[299] Christopher G Small. A survey of multidimensional medians. International Statistical Review/Revue Internationale de Statistique, 58(3):263–277, 1990.

[300] Stephen GZ Smith and István Tomon. The poset on connected graphs is sperner. Journal of Combinatorial Theory, Series A, 150:162–181, 2017.

[301] Stephen P Smith and Anil K Jain. Testing for uniformity in multidimensional data. IEEE transactions on pattern analysis and machine intelligence, PAMI-6(1):73–81, 1984.

[302] Michael Smithson. Confidence intervals. Number 140 in Quantitative Applications in the Social Sciences. Sage, London, UK, 2003.

[303] Miguel A Sordo, Marilia C de Souza, and Alfonso Suárez-Llorens. Testing variability orderings by using gini's mean differences. Statistical Methodology, 32:63–76, 2016.

[304] Jana Špirková and Juliana Beganová. Aggregation using penalty and moderate deviation functions. In Conference of the European Society for Fuzzy Logic and Technology, pages 247–258. Springer, 2023.

[305] J Michael Steele. Growth rates of euclidean minimal spanning trees with power weighted edges. The Annals of Probability, 16(4):1767–1787, 1988.

[306] Michael A Stephens. Asymptotic calculations of functions of expected values and covariances of order statistics. Canadian Journal of Statistics, 18(3):265–270, 1990.

[307] Stephen M Stigler. Linear functions of order statistics. The Annals of Mathematical Statistics, 40(3):770–788, 1969.

[308] Pan Su, Tianhua Chen, Haoyu Mao, Jianyang Xie, Yitian Zhao, and Jiang Liu. On the application of preaggregation functions to fuzzy pattern tree. In 2019 IEEE International Conference on Fuzzy Systems, pages 1–6. IEEE, 2019.

[309] Yixiang Sun and Nana Geng. Forecasting civil aviation incident rate in china using a combined prediction model. Journal of advanced transportation, 2021(1):5521526, 2021.

[310] Yunxin Sun, Ke Li, Yi Yan, Xinguo Wei, and Chenghui Zhang. Combination load forecasting method for cchp system based on iowa operator. In 2017 Chinese Automation Congress, pages 4193–4197. IEEE, 2017.

[311] Zdenko Takáč. Aggregation of fuzzy truth values. Information Sciences, 271:1–13, 2014.

[312] Brian S Thomson. Real functions, volume 1170. Springer, New York, USA, 2006.

[313] Yiu Kuen Tse and Xibin Zhang. A monte carlo investigation of some tests for stochastic dominance. Journal of statistical computation and simulation, 74(5):361–378, 2004.

[314] Jasper J Van De Gronde and Jos BTM Roerdink. Frames, the loewner order and eigendecomposition for morphological operators on tensor fields. Pattern Recognition Letters, 47:40–49, 2014.

[315] Aad W van der Vaart, Jon A Wellner, Aad W van der Vaart, and Jon A Wellner. M-estimators. In Weak Convergence and Empirical Processes: With Applications to Statistics, pages 284–308. Springer, New York, USA, 1996.

[316] Vladimir V'yugin and Vladimir Trunov. Online aggregation of probability forecasts with confidence. Pattern Recognition, 121:108193, 2022.

[317] Andre Wakefield. Pierre simon laplace, 1749-1827: A determined scientist. Technology and Culture, 48(1):221–222, 2007.

[318] Helen M Walker. The contributions of karl pearson. Journal of the American Statistical Association, 53(281):11–22, 1958.

[319] David S Watkins. Fundamentals of matrix computations. John Wiley & Sons, New Jersey, USA, 2004.

[320] Jorg D Wichard and Maciej Ogorzalek. Time series prediction with ensemble models. In 2004 IEEE international joint conference on neural networks, volume 2, pages 1625–1630. IEEE, 2004.

[321] Frank Wilcoxon. Individual comparisons by ranking methods. In Breakthroughs in Statistics: Methodology and Distribution, pages 196–202. Springer, 1992.

[322] Tim Wilkin and Gleb Beliakov. Weakly monotonic averaging functions. International Journal of Intelligent Systems, 30(2):144–169, 2015.

[323] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. Biometrika, 20A(1/2):32–52, 1928.

[324] Eugene Wong and Moshe Zakai. Markov processes on the plane. Stochastics: An International Journal of Probability and Stochastic Processes, 15(4):311–333, 1985.

[325] Jian-Zhang Wu, Gleb Beliakov, Simon James, and Marek Gagolewski. Random generation of linearly constrained fuzzy measures and domain coverage performance evaluation. Information Sciences, 659:120080, 2024.

[326] Zeshui Xu and Ronald R Yager. Some geometric aggregation operators based on intuitionistic fuzzy sets. International journal of general systems, 35(4):417–433, 2006.

[327] Ronald R Yager. Families of owa operators. Fuzzy Sets and Systems, 59(2):125–148, 1993.

[328] Ronald R Yager. Induced aggregation operators. Fuzzy Sets and Systems, 137(1):59–69, 2003.

[329] Ronald R Yager. Centered owa operators. Soft Computing, 11(7):631–639, 2007.

[330] Ronald R Yager. Evaluating choquet integrals whose arguments are probability distributions. IEEE Transactions on Fuzzy Systems, 24(4):957–965, 2015.

[331] Ronald R Yager. On the owa aggregation with probabilistic inputs. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 23(Suppl. 1):143–162, 2015.

[332] Ka-Veng Yuen, Peng-Fei Liang, and Sin-Chi Kuok. Online estimation of noise parameters for kalman filter. Structural Engineering and Mechanics, 47(3):361–381, 2013.

[333] V Zardasht. A test for the increasing convex order based on the cumulative residual entropy. Journal of the Korean Statistical Society, 44:491–497, 2015.

[334] Yiying Zhang and Ka Chun Cheung. On the increasing convex order of generalized aggregation of dependent random variables. Insurance: Mathematics and Economics, 92:61–69, 2020.

[335] Hua Zhao, Zeshui Xu, Mingfang Ni, and Shousheng Liu. Generalized aggregation operators for intuitionistic fuzzy sets. International Journal of Intelligent Systems, 25(1):1–30, 2010.

[336] Xi Zhao, Alfred Stein, and Xiaoling Chen. Application of random sets to model uncertainties of natural entities extracted from remote sensing images. Stochastic Environmental Research and Risk Assessment, 24:713–723, 2010.

[337] Jiaming Zhu, Peng Wu, Huayou Chen, Jinpei Liu, and Ligang Zhou. Carbon price forecasting with variational mode decomposition and optimal combined model. Physica A: Statistical Mechanics and Its Applications, 519:140–158, 2019.

[338] Jiaming Zhu, Peng Wu, Huayou Chen, Ligang Zhou, and Zhifu Tao. A hybrid forecasting approach to air quality time series based on endpoint condition and combined forecasting model. International Journal of Environmental Research and Public Health, 15(9):1941, 2018.

[339] Nicolás Zumelzu, Benjamín Bedregal, Edmundo Mansilla, Humberto Bustince, and Roberto Díaz. Admissible orders on fuzzy numbers. IEEE Transactions on Fuzzy Systems, 30(11):4788–4799, 2022.