



Universidad de Oviedo
Programa de Doctorado
en Matemáticas y Estadística

Comparison of alternatives under uncertainty and imprecision

Tesis Doctoral

Ignacio Montes Gutiérrez



puede obtener mediante el Teorema de Sklar. Por ello, resulta necesario investigar una versión imprecisa de este resultado, que tendrá importantes aplicaciones en los órdenes estocásticos bivariantes definidos bajo imprecisión.

Si las alternativas se definen bajo imprecisión, pero no bajo incertidumbre, éstas se modelarán mediante conjuntos intuicionísticos. Para su comparación se introduce una teoría matemática de comparación de este tipo de conjuntos, dando especial relevancia al concepto de IF-divergencia. Estas medidas de comparación de conjuntos intuicionísticos poseen numerosas aplicaciones, como pueden ser en el reconocimiento de patrones o la teoría de la decisión. Los conjuntos intuicionísticos permiten grados de pertenencia y de no pertenencia, y por ello resultan un buen modelo bipolar. Dado que las probabilidades imprecisas también son utilizadas en el contexto de la información bipolar, se estudiarán las conexiones entre ambas teorías. Estos resultados mostrarán tener interesantes aplicaciones, y en particular permitirán extender la dominancia estocástica para la comparación de más de dos p-boxes.

RESUMEN (en Inglés)

In real life situations it is common to deal with the comparison of alternatives. The alternatives to be compared are sometimes defined under some lack of information. Two lacks of information are considered: uncertainty and imprecision. Uncertainty refers to situations in which the possible results of the experiment are precisely described, but the exact result of the experiment is unknown; imprecision refers to situations in which the result of the experiment is known but it cannot be precisely described. In this work, uncertainty is modelled by means of Probability Theory, imprecision is modelled by means of IF-set Theory, and the Theory of Imprecise Probabilities is used when both lacks of information hold together.

Alternatives under uncertainty are modelled by means of random variables. Thus, a stochastic order is needed for their comparison. In this work two particular stochastic orders are considered: stochastic dominance and statistical preference. The former is one of the most usual methods used in the literature and the latter is the most adequate method for comparing qualitative variables. Some properties about such methods are investigated. In particular, although stochastic dominance is related to the expectation of some transformation of the random variables, statistical preference is related to a different location parameter: the median. In addition, some conditions, related to the copula that links the random variables, under which stochastic dominance and statistical preference are related are given. Both stochastic orders are defined for the pairwise comparison of random variables. Thus, an extension of statistical preference for the comparison of more than two random variables is defined, and its main properties are studied.

When the alternatives are defined under uncertainty and imprecision, each one is represented by a set of random variables. For comparing them, stochastic orders are extended for the comparison of sets of random variables instead of single ones. When the stochastic order is either stochastic dominance or statistical preference, the comparison of sets of random variables can be related to the comparison of elements of the imprecise probability theory, like p-boxes. Two particular instances of comparison of sets of random variables, common in decision making problems, are studied: the comparison of random variables with imprecision on the utilities or in the beliefs. The former situation is modelled by random sets, and then their sets of measurable selections are compared, and the second is modelled by a set of probabilities. When there is imprecision in the marginal distributions of the random variables, the joint distribution cannot be obtained from Sklar's Theorem. For this reason, an imprecise version of Sklar's Theorem is given, and its applications to bivariate stochastic orders under imprecision are showed.

Alternatives defined under imprecision, but not under uncertainty, are modeled by means of IF-sets. For their comparison a mathematical theory of comparison of IF-sets is given, focusing on



Vicerrectorado de Internacionalización
y Postgrado
Universidad de Oviedo



a particular type of measure called IF-divergences. This measure has several applications, like for instance in pattern recognition or decision making. IF-sets are used to model bipolar information because they allow membership and non-membership degrees. Since imprecise probabilities also allow to model bipolarity, a connection between both theories is established. As an application of this connection, an extension of stochastic dominance for the comparison of more than two p-boxes is showed.

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

Contents

Agradecimientos	xv
Resumen	xvii
Abstract	xix
1 Introduction	1
2 Basic concepts	9
2.1 Stochastic orders	9
2.1.1 Stochastic dominance	10
2.1.2 Statistical preference	15
2.2 Imprecise probabilities	24
2.2.1 Coherent lower previsions	24
2.2.2 Conditional lower previsions	27
2.2.3 Non-additive measures	29
2.2.4 Random sets	31
2.3 Intuitionistic fuzzy sets	34
3 Comparison of alternatives under uncertainty	43
3.1 Properties of the statistical preference	44

3.1.1	Basic properties and intuitive interpretation of the statistical preference	44
3.1.2	Characterizations of statistical preference	52
3.2	Relationship between stochastic dominance and statistical preference	77
3.2.1	Independent random variables	78
3.2.2	Continuous comonotonic and countermonotonic random variables	80
3.2.3	Discrete comonotonic and countermonotonic random variables with finite supports	83
3.2.4	Random variables coupled by an Archimedean copula	86
3.2.5	Other relationships between stochastic dominance and statistical preference	90
3.2.6	Examples on the usual distributions	93
3.3	Comparison of n variables by means of the statistical preference	102
3.3.1	generalisation of the statistical preference	104
3.3.2	Basic properties	106
3.3.3	Stochastic dominance Vs general statistical preference	111
3.3.4	General statistical preference Vs n^{th} degree stochastic dominance	127
3.4	Applications	129
3.4.1	Comparison of fitness values	129
3.4.2	General statistical preference as a tool for linguistic decision making	141
3.5	Conclusions	145
4	Comparison of alternatives under uncertainty and imprecision	147
4.1	generalisation of the binary relations	148
4.1.1	Imprecise stochastic dominance	157
4.1.2	Imprecise statistical preference	193
4.2	Modelling imprecision in decision making problems	202

4.2.1	Imprecision on the utilities	202
4.2.2	Imprecision on the beliefs	206
4.3	Modelling the joint distribution	209
4.3.1	Bivariate distribution with imprecision	210
4.3.2	Imprecise copulas	228
4.3.3	The role of imprecise copulas in the imprecise orders	237
4.4	Applications	243
4.4.1	Comparison of Lorenz curves	243
4.4.2	Comparison of cancer survival rates	247
4.4.3	Multiattribute decision making	249
4.5	Conclusions	253
5	Comparison of alternatives under imprecision	257
5.1	Measures of comparison of IF-sets	258
5.1.1	Comparison of IF-sets	267
5.1.2	Properties of the IF-divergences	273
5.1.3	Examples of IF-divergences and IF-dissimilarities	279
5.1.4	Local IF-divergences	290
5.1.5	IF-divergences Vs Divergences	314
5.2	Connecting IVF-sets and imprecise probabilities	330
5.2.1	Probabilistic information of IVF-sets	331
5.2.2	Connection with other approaches	346
5.3	Applications	349
5.3.1	Application to pattern recognition	350
5.3.2	Application to decision making	352
5.3.3	Using IF-divergences to extend stochastic dominance	358
5.4	Conclusions	362

Conclusiones y trabajo futuro	363
Concluding remarks	369
A Appendix: Basic Results	377
List of symbols	385
Alphabetic index	390
List of Figures	392
List of Tables	395
Bibliography	397

Agradecimientos

Esta memoria abarca el trabajo realizado desde que en Julio de 2009 finalicé la licenciatura en Matemáticas. Por aquel entonces recibí el apoyo de varios miembros del Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática para realizar una tesis doctoral. De entre ellos, debo destacar los consejos que siempre me dieron los Profesores Pedro Gil y Santos Domínguez.

La oportunidad de comenzar este trabajo me surgió cuando Susana Montes se puso en contacto conmigo para ofrecerme la posibilidad de adentrarme en el apasionante mundo de la investigación, y es de justicia calificar este momento como uno de los más importantes de toda mi vida. Desde entonces me encuentro en el mejor ambiente posible de trabajo, arropado en todo momento por los miembros del proyecto UNIMODE: Susana Díaz, Davide Martinetti, Enrique Miranda y la propia Susana Montes.

Es de justicia nombrar de manera especial a mis directores Susana Montes y Enrique Miranda (y aunque no figure como directora, también a Susana Díaz). Este trabajo ha sido posible gracias a ellos, a su dedicación, a su paciencia, a su pasión por las matemáticas, a sus ganas de hacer las cosas bien. Podría rellenar muchas páginas explicando lo agradecido que les estoy y lo afortunado que me siento por pertenecer a este maravilloso grupo de trabajo, pero sinceramente mis palabras nunca llegarían a reflejar mis sentimientos.

A lo largo de estos años ha habido muchas personas a las que les debo mis agradecimientos, comenzando por mis compañeros de carrera y los miembros del Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática. Al personal de la UCE, y en particular a Tania y Patri, por su incondicional apoyo y su sincera amistad. A Gert de Cooman y todo su equipo, y a Paolo Vicig y Renato Pelessoni debo agradecerles su amabilidad durante las estancias que realicé en las Universidades de Gante y de Trieste. A los miembros de la comunidad científica nacional e internacional por su amabilidad y buenos consejos. Y en definitiva, a todos aquellos que de una u otra manera han colaborado o aportado en mi formación.

Por otra parte, este trabajo no habría sido posible sin la financiación, a través de una

Beca de Formación de Profesorado Universitario, del Ministerio de Educación, así como a la Ayuda recibida por parte del Vicerrectorado de Investigación y Campus de Excelencia de la Universidad de Oviedo que me permitió realizar una estancia de investigación en la Universidad de Trieste.

Y por supuesto, darles las gracias a mi familia y a Almudena . . . simplemente por todo.

Resumen

En muchas situaciones de la vida real es necesario comparar alternativas. Además, es habitual que estas alternativas estén definidas bajo falta de información. En esta memoria se consideran dos tipos de falta de información: incertidumbre e imprecisión. La incertidumbre se refiere a situaciones en las cuales los posibles resultados del experimento son conocidos y se pueden describir completamente, pero el resultado del mismo no es conocido; mientras que en las situaciones bajo imprecisión, se conoce el resultado del experimento, pero no es posible describirlo con precisión. Por tanto, la incertidumbre se modelará mediante la Teoría de la Probabilidad, mientras que la imprecisión será modelada mediante la Teoría de los Conjuntos Intuicionísticos. Además, cuando ambas faltas de información aparezcan simultáneamente, se utilizará la Teoría de las Probabilidades Imprecisas.

Cuando las alternativas a comparar estén definidas bajo incertidumbre, éstas se modelarán mediante variables aleatorias. Por tanto, para compararlas será necesario utilizar un orden estocástico. En esta memoria se consideran dos órdenes: la dominancia estocástica y la preferencia estadística. El primero de ellos es uno de los métodos más utilizados en la literatura, mientras que el segundo es el método óptimo de comparación de variables cualitativas. Para estos métodos se han estudiado varias propiedades. En particular, si bien es conocido que la dominancia estocástica está relacionada con la comparación de las esperanzas de determinadas transformaciones de las variables, se prueba que la preferencia estadística está más ligada a otro parámetro de localización, la mediana. Además, se han encontrado situaciones bajo las cuales la dominancia estocástica está relacionada con la preferencia estadística. Estos dos órdenes estocásticos han sido definidos para comparar variables aleatorias por pares. Por esta razón se ha definido una extensión de la preferencia estadística para la comparación simultánea de más de dos variables y se han estudiado varias propiedades.

Cuando las alternativas están definidas en un marco de incertidumbre e imprecisión, cada una de ellas se modelará mediante un conjunto de variables aleatorias. Dado que los órdenes estocásticos comparan variables aleatorias, es necesario realizar su extensión para la comparación de conjuntos de variables. Cuando el orden estocástico utilizado es la dominancia estocástica o la preferencia estadística, la comparación de los conjuntos de

variables aleatorias está claramente relacionada con la comparación de elementos propios de la teoría de las probabilidades imprecisas, como pueden ser las p -boxes. Gracias al modelo general que desarrollaremos, se podrán estudiar en particular dos situaciones habituales en los problemas de la teoría de la decisión: la comparación de variables aleatorias bajo utilidades o bajo creencias imprecisas. El primer problema se modelará mediante conjuntos aleatorios, y por lo tanto su comparación se realizará a través de sus conjuntos de selecciones medibles. El segundo problema será modelado mediante un conjunto de probabilidades. Cuando las distribuciones marginales de las variables están definidas bajo imprecisión, la distribución conjunta no se puede obtener mediante el Teorema de Sklar. Por ello, resulta necesario investigar una versión imprecisa de este resultado, que tendrá importantes aplicaciones en los órdenes estocásticos bivariantes definidos bajo imprecisión.

Si las alternativas se definen bajo imprecisión, pero no bajo incertidumbre, éstas se modelarán mediante conjuntos intuicionísticos. Para su comparación se introduce una teoría matemática de comparación de este tipo de conjuntos, dando especial relevancia al concepto de IF-divergencia. Estas medidas de comparación de conjuntos intuicionísticos poseen numerosas aplicaciones, como pueden ser en el reconocimiento de patrones o la teoría de la decisión. Los conjuntos intuicionísticos permiten grados de pertenencia y de no pertenencia, y por ello resultan un buen modelo bipolar. Dado que las probabilidades imprecisas también son utilizadas en el contexto de la información bipolar, se estudiarán las conexiones entre ambas teorías. Estos resultados mostrarán tener interesantes aplicaciones, y en particular permitirán extender la dominancia estocástica para la comparación de más de dos p -boxes.

Abstract

In real life situations it is common to deal with the comparison of alternatives. The alternatives to be compared are sometimes defined under some lack of information. Two lacks of information are considered: uncertainty and imprecision. Uncertainty refers to situations in which the possible results of the experiment are precisely described, but the exact result of the experiment is unknown; imprecision refers to situations in which the result of the experiment is known but it cannot be precisely described. In this work, uncertainty is modelled by means of Probability Theory, imprecision is modelled by means of IF-set Theory, and the Theory of Imprecise Probabilities is used when both lacks of information hold together.

Alternatives under uncertainty are modelled by means of random variables. Thus, a stochastic order is needed for their comparison. In this work two particular stochastic orders are considered: stochastic dominance and statistical preference. The former is one of the most usual methods used in the literature and the latter is the most adequate method for comparing qualitative variables. Some properties about such methods are investigated. In particular, although stochastic dominance is related to the expectation of some transformation of the random variables, statistical preference is related to a different location parameter: the median. In addition, some conditions, related to the copula that links the random variables, under which stochastic dominance and statistical preference are related are given. Both stochastic orders are defined for the pairwise comparison of random variables. Thus, an extension of statistical preference for the comparison of more than two random variables is defined, and its main properties are studied.

When the alternatives are defined under uncertainty and imprecision, each one is represented by a set of random variables. For comparing them, stochastic orders are extended for the comparison of sets of random variables instead of single ones. When the stochastic order is either stochastic dominance or statistical preference, the comparison of sets of random variables can be related to the comparison of elements of the imprecise probability theory, like p-boxes. Two particular instances of comparison of sets of random variables, common in decision making problems, are studied: the comparison of random variables with imprecision on the utilities or in the beliefs. The former situation is modelled by random sets, and then their sets of measurable selections are compared,

and the second is modelled by a set of probabilities. When there is imprecision in the marginal distributions of the random variables, the joint distribution cannot be obtained from Sklar's Theorem. For this reason, an imprecise version of Sklar's Theorem is given, and its applications to bivariate stochastic orders under imprecision are showed.

Alternatives defined under imprecision, but not under uncertainty, are modeled by means of IF-sets. For their comparison a mathematical theory of comparison of IF-sets is given, focusing on a particular type of measure called IF-divergences. This measure has several applications, like for instance in pattern recognition or decision making. IF-sets are used to model bipolar information because they allow membership and non-membership degrees. Since imprecise probabilities also allow to model bipolarity, a connection between both theories is established. As an application of this connection, an extension of stochastic dominance for the comparison of more than two p-boxes is showed.

1 Introduction

The mathematical modeling of real life experiments can be rendered difficult by the presence of two types of lack of information: uncertainty and imprecision. We speak about uncertainty when the variables involved in the experiment are precisely described but we cannot predict beforehand the outcome of the experiment. This lack of information is usually modelled by means of Probability Theory. On the other hand, imprecision refers to situations in which the result of the experiment is known but it cannot be precisely described. One possible model for this situation is given by Fuzzy Set Theory or any of its extensions, such as the Theory of Intuitionistic Fuzzy Sets or the Theory of Interval-Valued Fuzzy Sets. Of course, there are also situations in which both uncertainty and imprecision appear together. In such cases, we can either combine probability theory and fuzzy sets, or consider the Theory of Imprecise Probabilities.

Fuzzy sets were introduced by Zadeh ([214]) as a more flexible model than crisp sets, which is particularly useful when dealing with linguistic information. A fuzzy set assigns a value to each element on the universe, called membership degree, which is interpreted as the degree in which the element fulfills the characteristic described by the set. Of course, crisp sets are particular cases of fuzzy sets, since every element either belongs (i.e., has membership degree 1) or does not (membership degree equals 0) to the set. Since their introduction, fuzzy sets have become a very popular research topic, and nowadays several international journals, conferences and societies are devoted to them. For a complete study on fuzzy sets, we remit the reader to some usual references like ([71, 101]).

In 1983, Atannasov ([4]) proposed a generalization of fuzzy sets, called the theory of Intuitionistic Fuzzy Sets (IF-sets, for short). In the subsequent years he continued developing his idea ([5, 7]), and now it has become a commonly accepted generalization of fuzzy sets. While fuzzy sets give a degree of membership of every element to the set, an IF-set assigns both a degree of membership and a degree of non-membership of any element to the set, with the natural restriction of that their sum must not exceed 1. Every IF-set has a degree of indeterminacy or uncertainty, that is, one minus the sum of the degrees of membership and non-membership. In this sense we can see that every fuzzy set is in particular an IF-set, since the non-membership degree of the fuzzy set is

one minus its membership degree: the indeterminacy degree of a fuzzy set equals zero. For this reason IF-sets have become a very useful tool in order to model situations in which human answers are present: *yes*, *no* or *does not apply*, like for example human votes ([8]). On the other hand, Zadeh also proposed several generalizations of fuzzy sets ([216]). In particular, he introduced interval-valued fuzzy sets (IVF-sets, for short): when the membership degree of an element to the set cannot be precisely determined, it assigns an interval that contains the real membership degree. Although IF-sets and IVF-sets differ on the interpretation, they are formally equivalent (see [30]). These theories have been applied to different areas, like decision making ([194]), logic programming ([9, 10]), medical diagnosis ([48]), pattern recognition ([92]) and interesting theoretical developments are still being made (see for example [68, 97, 120]).

The second pillar of this dissertation is the theory of Imprecise Probabilities. Imprecise Probability is a generic term that refers to all mathematical models that serve as an alternative and a generalization to probability models in cases of imprecise knowledge. It includes possibility measures ([217]), Choquet capacities ([39]), belief functions ([187]) or coherent lower previsions ([205]), among others. One model that will be of particular interest for us is that of p-boxes. A p-box ([75]) is determined by an ordered pair of functions called lower and upper distribution functions, and it is given by all the distribution functions bounded between them. Troffaes et al. ([198, 201]) have investigated the connection between p-boxes and coherent lower probabilities ([205]). In particular, they found conditions under which a p-box defines a coherent lower probability. In some recent papers ([64, 65, 199, 200]), the authors have explored the connection between p-boxes and other usual models included in the theory of imprecise probabilities, such as possibilities, belief functions or clouds ([168]), among others.

This memory deals with the comparison of alternatives under lack of information. As we mentioned before, we shall consider the comparison under uncertainty, imprecision or both. On the one hand, alternatives under uncertainty are modelled by means of random variables. Random variables are one tool of the probability theory that provide a formal background to model non-deterministic situations, that is, situations where randomness is present. The comparison of random variables is a long standing problem that has been tackled from many points of view (see among others [18, 90, 98, 106, 188, 192, 210]). Its practical interest is clear since many real life processes are modelled by random variables. The procedures of comparison are referred to as stochastic orders. Indeed, stochastic ordering is a very popular topic within Economics ([11, 109]), Finance ([110, 173]), Social Welfare ([77]), Agriculture ([95]), Soft Computing ([180, 183]) or Operational Research ([171]), among others.

One classical way of pairwise ordering random variables is stochastic dominance ([108, 208]), a generalization of the expected utility model. First degree stochastic dominance, that seems to be the most widely used method, orders random variables by comparing their cumulative distribution functions (or their survival functions). Its main drawback is that it imposes a very strong condition to get an order, so many pairs

of random variables are deemed incomparable. Because of this fact, a second definition, called second degree stochastic dominance is also used, specially in Economics ([98, 139]). Although less restrictive, it still does not establish a complete order between random variables. In fact, we can weaken progressively the notion of stochastic dominance, and talk of stochastic dominance of n -th order.

One interesting alternative stochastic order is statistical preference, particularly when comparing qualitative random variables, taking into account the results by Dubois et al. ([67]). Although it was introduced by De Schuymer et al. ([55, 57]), it is possible to find similar methods in the literature (see [25, 26, 210]). The notion of statistical preference is based on a probabilistic relation, also called reciprocal relation ([21]), that measures the degree of preference of one random variable over the other one. Furthermore, since statistical preference depends on the joint distribution of the random variables, it depends on the copula ([166]) that links them. Recall that from Sklar's Theorem ([189]) it is known that for any two random variables there exists a function, called copula, that allows to express the joint cumulative distribution function in terms of the marginals. Then, statistical preference depends on such copula. The main drawback of this method is its lack of transitivity. Some authors have been investigating which kind of transitivity properties are satisfied by statistical preference, and in particular they focused on cycle-transitivity (see [14, 15, 16, 49, 54, 56, 58, 121, 122]).

When the alternatives to be compared are defined under both uncertainty and imprecision, the problem of comparing sets of random variables arises. Here we understand the set of random variables from an epistemic point of view: we assume that the set of random variables contains the true random variable, but such random variable is unknown ([73]). This situation is not uncommon in decision making under uncertainty, where there is vague or conflicting information about the probabilities or the utilities associated to the different alternatives. We may think for instance of conflicts among the opinions of several experts, limits or errors in the observational process, or simply partial or total ignorance about the process underlying the alternatives. In any of such cases, the elicitation of an unique probability/utility model for each of the alternatives may be difficult and its use, questionable.

Indeed, one of the solutions that have been proposed for situations like this is to consider a robust approach, by means of a set of probabilities and utilities. The use of this approach to compare two alternatives is formally equivalent to the comparison of two sets of alternatives, those associated to each possible probability-utility pair. Hence, it becomes useful to consider comparison methods that allow us to deal with sets of alternatives instead of single ones.

However, the way to compare of sets of alternatives is no longer immediate: we may compare all possibilities within each of the sets, or also select some particular elements of each set, to take into account phenomena of risk aversion, for instance. This gives rise to a number of possibilities. Moreover, even in the simpler case where we choose one alternative from each set, we must still decide which criterion we shall consider to

determine the preferred one. There is quite an extensive literature on how to deal with imprecise beliefs and utilities when our choice is made by means of an expected utility model ([12, 165, 178, 186]). However, the problem has almost remained unexplored for other choice functions. For this reason, we shall extend stochastic orders for the comparison of sets of random variables, and we shall see that the proposed extension is connected to the imprecise probability theory.

The last situation to be studied is the comparison of alternatives under imprecision but without uncertainty. In this case the alternatives will be described by means of IF-sets. Within fuzzy set theory, several types of measures of comparison have been defined, with the goal of quantifying how different two fuzzy sets are. The more usual measures of comparison are dissimilarities ([119]), dissimilitudes ([44]) and divergences ([159]). Other authors, like Bouchon-Meunier et al. ([27]), defined a general axiomatic framework for the comparison of fuzzy sets, that include the aforementioned measures as particular cases. Montes ([159]) made a complete study of the divergences as a measure of comparison of fuzzy sets. In particular, she introduced a particular kind of divergences, called local divergences, that have proven to be very useful.

Distances between fuzzy sets are also important for many practical applications. For instance, Bhandari et al. ([22]) proposed a divergence measure for fuzzy sets inspired by the notion of divergence between two probability distributions, and used this fuzzy divergence measure in the framework of image segmentation. Several other attempts within the same field have been considered ([23, 34, 74]). For instance, the fuzzy divergence measure of Fan and Xie is based (unlike the proposal of Bhandari and Pal) on the exponential entropy of Pal and Pal ([175]); the same spirit is followed in [34].

However, in the framework of IF-sets only the notion of distance as well as several examples of IF-dissimilarities have been given (see for example [36, 37, 85, 89, 92, 111, 113, 114, 138, 193]). Nevertheless, the need for a formal mathematical theory of comparison of IF-sets still persists.

Furthermore, IF-sets are a very useful tool to represent bipolar information: the membership and non-membership degree of every element to the set. Since bipolar models are also being studied within the framework of imprecise probabilities (see for instance [64, 65, 72, 73]), it becomes natural to investigate the connection between both approaches to the modeling of bipolar information.

The rest of the work is organized as follows. Chapter 2 introduces the basic notions that will be necessary along the work. In the first part we deal with stochastic orders, focusing on stochastic dominance, that is based on the comparison of the cumulative distribution functions of the random variables, and statistical preference, that is based on a probabilistic relation and makes use of the joint distribution. In order to express this joint distribution as a function of the marginals, we need to introduce some notions of the theory of copulas. Then, we make a brief introduction to the theory of imprecise probabilities. On the first part we define coherent lower previsions and we recall the

basic results we shall use later on. Then, we focus on particular cases of coherent lower probabilities: n -monotone capacities, belief functions, possibility measures and clouds. We also define random sets and show their connections with imprecise probability theory. Finally, we make an overview of IF-sets theory. First, we explain the semantic differences between IF-sets and IVF-sets and show that both theories are formally equivalent. Then, we introduce the basic operations between these sets.

In Chapter 3 we investigate the comparison of alternatives under uncertainty, that will be modelled by means of random variables. Although some stochastic orders like stochastic dominance have already been widely explored in the literature, this is not the case for statistical preference. For this reason, we devote Section 3.1 to investigate the main properties of this relation, and we compare them to the ones of stochastic dominance ([149, 154]). While stochastic dominance has a well-known characterization in terms of the comparison of the expectations of adequate transformations of the random variables, there is not a characterization of statistical preference. For this aim, we investigate a possible characterization in terms of expectations ([150, 153]) and in terms of a different location parameter: the median ([148, 163]).

Although statistical preference and stochastic dominance are not related in general, in Section 3.2 we look for conditions under which first degree stochastic dominance implies statistical preference ([150]). Obviously, since statistical preference depends on the copula that links the variables, these conditions are related to such copula. Furthermore, we find that in some of the usual probability distributions, like Bernoulli, uniform, normal, etc, both stochastic dominance and statistical preference are equivalent for independent random variables ([151]).

We have already mentioned the lack of transitivity of statistical preference, which renders it unsuitable for comparing more than two random variables. In order to overcome this problem, we introduce in Section 3.3 an extension of statistical preference that preserves its philosophy and allows the comparison of more than two random variables ([140, 142]). We explore this new notion and give several properties that relate it to the classical notion of statistical preference. In order to illustrate the applicability of our results, Section 3.4.1 puts forward two different applications. We first use both stochastic dominance and statistical preference to compare fitness values associated to the output of genetic fuzzy systems ([143, 152, 162]), and then we use the generalization of statistical preference on a decision-making problem with linguistic variables.

In Chapter 4 we consider the comparison of alternatives under both uncertainty and imprecision. As we have already mentioned, in that case we model the alternatives by means of sets of random variables instead of single ones. We start in Section 4.1 by extending binary relations that are used to the comparison of random variables to the comparison of sets of random variables. This gives rise to six possible ways of comparing sets of random variables. In particular, we focus on the case where such binary relation is either stochastic dominance or statistical preference. We shall see that the use of stochastic dominance as binary relation is clearly connected to the comparison of the p-

boxes associated with the sets of random variables ([134, 155, 157]). We shall consider two particular cases in Section 4.2: the comparison of two random variables with imprecise utilities and the comparison of two random variables with imprecise beliefs ([156]). The former is modelled by means of random sets, and their comparison is made by means of their associated sets of measurable selections. In the latter, the imprecise beliefs are modelled by means of a set of probabilities in the initial space, instead of a single one. In this situation we can also define a set of random variables for each alternative. Then, both situations are particular cases of the more general situation studied in Section 4.1.

When there is imprecision about the probability of the initial space, the joint distribution of the random variables is also imprecisely determined. Because of this, it seems reasonable to investigate how the bivariate distribution, and in particular the bivariate cumulative distribution function, can be determined. We shall investigate the properties of bivariate p -boxes and how they can define a coherent lower probability ([135]). One particular instance where the joint distribution naturally arises is when dealing with copulas. Recall that copulas allow to determine the joint distribution function in terms of the marginals. However, when the marginal distribution functions are imprecisely described by means of p -boxes, it is unclear how to determine the joint distribution, and bivariate p -boxes prove to be a useful tool. In particular we show that, by considering an imprecise version of copulas it is possible to extend Sklar's Theorem to an imprecise framework ([176]).

Section 4.4 shows several applications of the results from Chapter 4. One possible application is the comparison of Lorenz Curves ([3, 11]), that represent the inequalities within countries/regions. Using our results, it is possible to compare sets of regions by means of stochastic dominance. Furthermore, imprecise stochastic dominance also allows to compare survival rates of different cancer grouped by sites. We conclude the chapter showing another application in decision making.

In Chapter 5 we investigate how to compare alternatives under imprecision. The alternatives are modelled by means of IF-sets, and we propose methods for comparing IF-sets. In Section 5.1 we recall the comparison measures that can be found in the literature: IF-dissimilarities and distances for IF-sets. We also introduce IF-divergences and IF-dissimilitudes ([141]). We investigate the relationships among these measures and we justify that our preference for IF-divergences in that they impose stronger conditions, avoiding thus counterintuitive examples ([145, 161]). We also try to define a general measure of comparison of IF-sets as done by Bouchon-Meunier et al. ([27]) for fuzzy sets. This allows us to define a general function that contains IF-dissimilarities, IF-divergences and distances as particular cases ([158]). Then we introduce a particular type of IF-divergences, that are those that satisfy a local property. We investigate their properties and give several examples ([147]). We conclude the section studying the connection between IF-divergences and divergences for fuzzy sets. In particular, we show how we can define IF-divergences from divergences for fuzzy sets and, conversely, how to build divergences for fuzzy sets from IF-divergences ([146]).

Since both imprecise probabilities and IF-sets are used to model bipolarity, we investigate in Section 5.2 the connection between both approaches. We establish that when IF-sets are defined in a probability space, they can be interpreted as random sets, and this allows to connect them with imprecise probabilities, since it is possible to define a credal set and a lower and upper probability. We investigate under which conditions the probabilistic information encoded by the credal set is the same than the one of the set of measurable selections. We also investigate the relationship between our approach and other works in the literature, like the one of Grzegorzewski and Mrowka ([86]).

We conclude the chapter showing several applications of the results. On the one hand we show how IF-divergences can be applied to decision making and pattern recognition. On the other hand, we explain how the connection between IF-sets and imprecise probabilities allows us to propose a generalization of stochastic dominance to the comparison of more than two p-boxes, and we illustrate our method comparing at the same time sets of Lorenz Curves.

We conclude this dissertation with some final remarks and a discussion of the most important future lines of research.

2 Basic concepts

In this chapter, we introduce the main notions that shall be employed in the rest of the work. We start by providing the definition of binary relations as comparison methods for random variables. Later, we consider the particular cases where the binary relation is either stochastic dominance or statistical preference, which are the two main stochastic orders we shall consider here.

Afterwards we make a brief introduction to Imprecise Probability theory, that shall be useful when we want to compare sets of random variables. To conclude the chapter, we recall the notion of intuitionistic fuzzy sets, that we shall use model situations where sets cannot be precisely described.

2.1 Stochastic orders

Stochastic orders are methods that determine a (total or partial) order on any given set of random quantities. Although several methods have been proposed in the last years (see for instance [139, 188]), here we shall focus on two particular cases: stochastic dominance and statistical preference. The former is possibly the most widespread method in the literature, and the latter is particularly useful when comparing qualitative variables, taking into account the axiomatization established by Dubois et al. ([67]).

Throughout, random variables are denoted by X, Y, Z, \dots , or X_1, X_2, \dots , and their associated cumulative distribution functions are denoted F_X, F_Y, F_Z, \dots , or F_{X_1}, F_{X_2}, \dots , respectively. We shall also assume that the random variables to be compared are defined on the same probability space.

Given two random variables X and Y defined from the probability space (Ω, \mathcal{A}, P) to an ordered space (Ω', \mathcal{A}') (which in most situations will be the set of real numbers), a binary relation \succeq is used to compare the variables. Then, $X \succeq Y$ means that X is at least as preferable as Y . This corresponds to a weak preference relation; from it a strict preference relation, indifference and also incomparable relation can also be defined:

Definition 2.1 Consider two random variables X and Y and a binary relation \succeq used to compare them.

- X is strictly preferred to Y with respect to \succeq , and is denoted by $X \succ Y$, if $X \succeq Y$ but $Y \not\succeq X$.
- X and Y are indifferent with respect to \succeq , and it is denoted by $X \equiv Y$, if $X \succeq Y$ and $Y \succeq X$.
- X and Y are incomparable with respect to \succeq , and it is denoted by $X \not\sim Y$, if $X \not\succeq Y$ and $Y \not\succeq X$.

Then, if \mathcal{D} denotes a set of random variables, according to [179], $(\mathcal{D}, \succeq, \equiv, \not\sim)$ forms a preference structure. In particular, if the relation \succeq is complete, that is, if there is not incomparability between the random variables, then $(\mathcal{D}, \succeq, \equiv)$ forms a preference structure without incomparable elements.

One instance of binary relation is the comparison of the expectations of the random variables, so that $X \succeq Y$ if and only if $E(X) \geq E(Y)$. This is also an example of a non-complete relation, because the comparison cannot be made when the expectation of the variable does not exist.

In the remainder of this section we introduce the definitions and notations that we shall use in the following chapters. Specifically, we consider the case in which the binary relation is either stochastic dominance or statistical preference. With respect to the first one, we recall the main types of stochastic dominance and some of its most important properties, such as its characterization by means of the comparison of the adequate expectations. Then, we provide an overview on statistical preference: we recall its definition and we also discuss briefly its main advantages as a stochastic order.

2.1.1 Stochastic dominance

Stochastic dominance is one of the most used methods for the pairwise comparison of random variables we can find in the literature. Besides to the usual economic interpretation (see [110]), this notion has also been applied in other frameworks such as Finance ([109]), Social Welfare ([11]), Agriculture ([95]) or Operations Research ([171]), among others. We next recall its definition and basic notions related to them, and also its main properties.

Stochastic dominance is a method based on the comparison of the cumulative distribution functions of the random variables.

Definition 2.2 Let X and Y be two real-valued random variables, and let F_X and F_Y denote their respective cumulative distribution functions. X stochastically dominates Y

by the first degree, or *simply stochastically dominates*, when no confusion is possible, and it is denoted by $X \succeq_{\text{FSD}} Y$, if it holds that

$$F_X(t) = P(X \leq t) \leq P(Y \leq t) = F_Y(t) \text{ for every } t \in \mathbb{R}. \quad (2.1)$$

One of the most important drawbacks of this definition is that (first degree) stochastic dominance is a non-complete relation, that is, it is possible to find random variables X and Y such that neither $X \succeq_{\text{FSD}} Y$ nor $Y \succeq_{\text{FSD}} X$, as we can see in the following example.

Example 2.3 Consider two random variables X and Y such that X follows a Bernoulli distribution with parameter 0.6 and Y takes a fixed value $c \in (0, 0.6)$ with probability 1. Then, there is not first degree stochastic dominance between them:

$$F_X(0) = 0.4 > 0 = F_Y(0) \text{ but } F_X(c) = 0.4 < 1 = F_Y(c). \blacklozenge$$

According to Definition 2.1, from this preference relation we can also define the strict stochastic dominance, the indifference and, as we have just seen, the incomparability relations:

- X stochastically dominates Y *strictly*, and denote it by $X \succ_{\text{FSD}} Y$, if and only if $F_X \leq F_Y$ and there is some $t \in [0, 1]$ such that $F_X(t) < F_Y(t)$.
- X and Y are *stochastically indifferent*, and denote it by $F_X \equiv_{\text{FSD}} F_Y$, if and only if they have the same distribution (usually denoted by $X \stackrel{d}{=} Y$).
- X and Y are *stochastically incomparable*, and denote it by $X \not\sim Y$, if there are t_1 and t_2 such that $F_X(t_1) > F_Y(t_2)$ and $F_Y(t_2) > F_X(t_2)$.

Remark 2.4 Here we have chosen the notation \succeq_{FSD} because it is the most frequent in the literature. However, (first degree) stochastic dominance has also been denoted by \succeq_1 , as in [55], or by \succeq_{st} , as in [188]. In that case, the authors used the name stochastic order instead of first degree stochastic dominance. \blacklozenge

As we see from its definition, (first degree) stochastic dominance only focuses on the marginal cumulative distribution functions, and its interpretation is the following: if $X \succeq_{\text{FSD}} Y$, then $F_X(t) \leq F_Y(t)$ for any t , or equivalently, $P(X > t) \geq P(Y > t)$ for any t . That is, we impose that at every point the probability of X to be greater than such point is greater than the probability of Y to be greater than the same point. Thus, X assigns greater probability to greater values. Figure 2.1 shows its graphical interpretation. Here, we can see how F_X is always below or at the same level than F_Y .

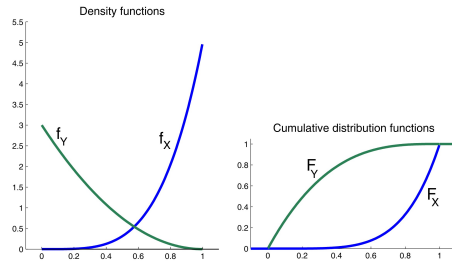


Figure 2.1: Example of first degree stochastic dominance: $X \succeq_{\text{FSD}} Y$

From an economic point of view, the interpretation is that the decision between the two random variables is rational, in the sense that for any threshold of profit the probability of going above this threshold is greater with the preferred variable ([110]).

The main drawback of this definition is that the inequality in Equation (2.1) is quite restrictive. There are many pairs of cumulative distribution functions that do not satisfy this inequality in any sense and therefore, the associated random variables cannot be ordered. This is the reason why we can consider other (weaker) degrees of stochastic dominance. Let us now introduce the second degree stochastic dominance.

Definition 2.5 Let X and Y be two real-valued random variables whose cumulative distribution functions are given by F_X and F_Y , respectively. X stochastically dominates Y by the second degree, and it is denoted by $X \succeq_{\text{SSD}} Y$, if it holds that:

$$\int_{-\infty}^t F_X(x)dx \leq \int_{-\infty}^t F_Y(y)dy \text{ for every } t \in \mathbb{R}. \quad (2.2)$$

As in Definition 2.2, we can also introduce the strict second degree stochastic dominance (\succ_{SSD}), the indifference (\equiv_{SSD}) and the incomparable ($\not\sim_{\text{SSD}}$) relations.

Note that, similar to Example 2.3, we can also see that incomparability is possible when dealing with second degree stochastic dominance.

Example 2.6 Consider the same random variables of Example 2.3. For these variables, the functions G_X^2 and G_Y^2 are defined by:

$$G_X^2(t) = \begin{cases} 0 & \text{if } t < 0. \\ 0.4t & \text{if } t \in [0, 1). \\ t - 0.6 & \text{if } t \geq 1. \end{cases} \quad G_Y^2(t) = \begin{cases} 0 & \text{if } t < c. \\ t - c & \text{if } t \geq c. \end{cases}$$

Then, X and Y are not ordered by means of the second degree stochastic dominance since:

$$G_X^2\left(\frac{c}{2}\right) = 0.2 \quad c > 0 = G_Y^2\left(\frac{c}{2}\right) \quad \text{but} \quad G_X^2(1) = 0.4 < 1 - c = G_Y^2(1),$$

since $c < 0.6$. ♦

Remark 2.7 Other authors (see for example [188]) call this method concave order, and they denote it by \geq_{cv} . It is also sometimes denoted by \succeq_2 ([55]). ♦

As we can see in Figure 2.2, when $X \succeq_{SSD} Y$, for any fixed t , the area below F_X until t is lower than the area below F_Y until t . This means that the X gathers more accumulated probability at greater points than Y .

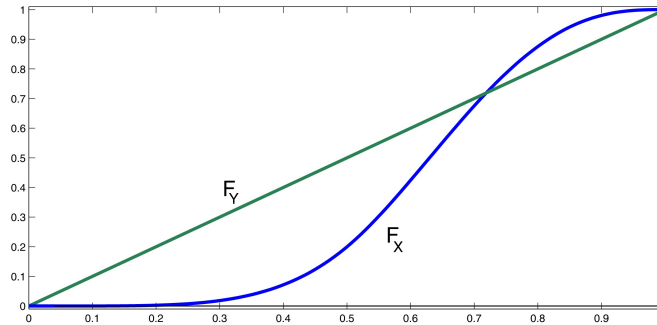


Figure 2.2: Example of second degree stochastic dominance: $X \succeq_{SSD} Y$.

From an economic point of view, second degree stochastic dominance means that the decision maker prefers the alternative that provides a bigger profit but also with less risk. That is, it is a rationality criterion under risk aversion (see [110]).

Similarly to Definitions 2.2 and 2.5, stochastic dominance can be defined for every degree n by relaxing the conditions in Equations (2.1) and (2.2).

Definition 2.8 Let X and Y be two real-valued random variables with cumulative distribution functions F_X and F_Y , respectively. X stochastically dominates Y by the n -th degree, for $n \geq 2$, and it is denoted by $X \succeq_{nSD} Y$, if it holds that:

$$G_X^n(t) = \int_{-\infty}^t G_X^{n-1}(x) d(x) \leq \int_{-\infty}^t G_Y^{n-1}(y) d(y) = G_Y^n(t) \quad \forall t \in \mathbb{R}, \quad (2.3)$$

where $G_X^1 = F_X$ and $G_Y^1 = F_Y$. In particular, this definition becomes the second degree stochastic dominance when $n = 2$.

Again, following the notation of Definition 2.1, we can introduce the strict n -th degree stochastic dominance (\succ_{nSD}), the indifference (\equiv_{nSD}) and the incomparability ($\not\sim_{nSD}$)

relations. Then, if \mathcal{D} denotes a set of random variables, $(\mathcal{D}, \succeq_{nSD}, \equiv_{nSD}, \not\prec_{nSD})$ forms a preference structure for any $n \geq 1$.

Clearly, first degree stochastic dominance imposes a stronger condition than second degree stochastic dominance, as we can see from Equations (2.1) and (2.2). Moreover, if we compare Equations (2.1) and (2.3) we deduce that first degree stochastic dominance is stronger than the n -th degree stochastic dominance for every n . Indeed, it is known that the n -th degree stochastic dominance is stronger than the m -th degree stochastic dominance for any $n < m$:

$$X \succeq_{nSD} Y \Rightarrow X \succeq_{mSD} Y \text{ for every } n < m, \quad (2.4)$$

while the converse does not hold in general.

Remark 2.9 *Stochastic dominance is a reflexive and transitive relation. However, since two different random variables may induce the same distribution, it is not antisymmetric. Moreover, as we have already noted, it is not complete because it allows incomparability. ♦*

One of the most important properties of stochastic dominance is its characterization by means of the expectation. Specifically, each of the types of stochastic dominance we have introduced can be characterized by the comparison of the expectations of adequate transformations of the variables considered.

Theorem 2.10 ([109, 139]) *Let X and Y be two random variables. For first and second degree stochastic dominance it holds that:*

- $X \succeq_{FSD} Y$ if and only if $E[u(X)] \geq E[u(Y)]$ for every increasing function $u : \mathbb{R} \rightarrow \mathbb{R}$.
- $X \succeq_{SSD} Y$ if and only if $E[u(X)] \geq E[u(Y)]$ for every increasing and concave function $u : \mathbb{R} \rightarrow \mathbb{R}$.

A function $u : \mathbb{R} \rightarrow \mathbb{R}$ is called n -monotone ([39]) if it is n -differentiable and for any $m \leq n$ and it fulfills $(-1)^{m+1}u^{(m)} \geq 0$. Then, if \mathcal{U}_n denotes the set of n -monotone functions, the following general equivalence holds:

$$X \succeq_{nSD} Y \Leftrightarrow E[u(X)] \geq E[u(Y)] \text{ for every } u \in \mathcal{U}_n. \quad (2.5)$$

In fact, from the proof of Theorem 2.10, it can be derived that:

$$X \succeq_{nSD} Y \Leftrightarrow E[u(X)] \geq E[u(Y)] \text{ for any } u \in \mathcal{U}_n^*, \quad (2.6)$$

where \mathcal{U}_n^* denote the set of n -monotone and bounded functions $u : \mathbb{R} \rightarrow \mathbb{R}$.

Equation (2.4) can also be derived from this result, since every n -monotone function is also m -monotone for any $m \leq n$.

Remark 2.11 *The characterization of the second degree stochastic dominance, based on the comparison of the mean of the concave and increasing functions, explains the nomenclature concave order mentioned in Remark 2.7.♦*

To conclude this paragraph, we list some interesting properties of first degree stochastic dominance that shall be useful in the next chapter.

Proposition 2.12 ([139, Theorem 1.2.13]) *If X and Y are real-valued random variables such that $X \succeq_{\text{FSD}} Y$ and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is an increasing function, then $\varphi(X) \succeq_{\text{FSD}} \varphi(Y)$.*

Proposition 2.13 ([139, Theorem 1.2.17]) *Let $\{X_i, Y_i : i = 1, \dots, n\}$ be independent and real-valued random variables. If $X_i \succeq_{\text{FSD}} Y_i$ for $i = 1, \dots, n$, then $X_1 + \dots + X_n \succeq_{\text{FSD}} Y_1 + \dots + Y_n$.*

Proposition 2.14 ([139, Theorem 1.2.14]) *Given the random variables $X, X_1, X_2, \dots, Y, Y_1, Y_2, \dots$ such that $X_n \xrightarrow{\mathcal{L}} X$ and $Y_n \xrightarrow{\mathcal{L}} Y$, if $X_n \succeq_{\text{FSD}} Y_n$ for every n , where $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution, then $X \succeq_{\text{FSD}} Y$.*

As a consequence of the previous result, first degree stochastic dominance is preserved by four kinds of convergence: distribution, probability, m^{th} -mean and almost sure.

For a more complete study on stochastic orders, we refer to [62, 109, 139, 188, 192].

2.1.2 Statistical preference

In the previous subsection we have mentioned that stochastic dominance is a pairwise comparison method that has been used in several areas, always with successful results. However, this method also presents some drawbacks: on the one hand, it is a non-complete crisp relation. This means that it is possible to find pairs of random variables such that n -th degree stochastic dominance does not order them for any n . Furthermore, stochastic dominance does not allow to establish degrees of preference. In fact, there are only three possibilities: either one random variable is preferred to the other, or they are indifferent or incomparable. In addition, it is a method with a high computational cost, since the n -th degree stochastic dominance requires the computation of $2(n-1)$ integrals.

These drawbacks made De Schuymer et al. ([55, 57]) introduce a new method for the pairwise comparison of the random variables, based on a probabilistic relation.

Definition 2.15 ([21]) *Given a set of alternatives \mathcal{D} , a probabilistic or reciprocal relation Q is a map $Q : \mathcal{D} \times \mathcal{D} \rightarrow [0, 1]$ such that $Q(a, b) + Q(b, a) = 1$ for any alternatives $a, b \in \mathcal{D}$.*

In our framework, the set of alternatives \mathcal{D} is considered to be made by random variables defined on the same probability space $(\Omega, \mathcal{P}(\Omega), P)$ to an ordered space (Ω', \mathcal{A}') . The probabilistic relation over \mathcal{D} is defined (see [55, Equation 3]) by:

$$Q(X, Y) = P(X > Y) + \frac{1}{2}P(X = Y), \quad (2.7)$$

where $(X, Y) \in \mathcal{D} \times \mathcal{D}$ and P denotes the joint probability of the bidimensional random vector (X, Y) . Clearly, Q is a probabilistic relation: it takes values in $[0, 1]$ and $Q(X, Y) + Q(Y, X) = 1$:

$$Q(X, Y) + Q(Y, X) = P(X > Y) + \frac{1}{2}P(X = Y) + \frac{1}{2}P(X = Y) + P(Y > X) = 1.$$

The above definition measures the preference degree of a random variable X over another random variable Y , in the sense that the greater the value of $Q(X, Y)$, the stronger the preference of X over Y . Hence, the closer the value $Q(X, Y)$ is to 1, the greater we consider X with respect to Y ; the closer $Q(X, Y)$ is to 0, the greater we consider Y to X ; and if $Q(X, Y)$ is around 0.5, both alternatives are considered indifferent. This fact can be seen in Figure 2.3.

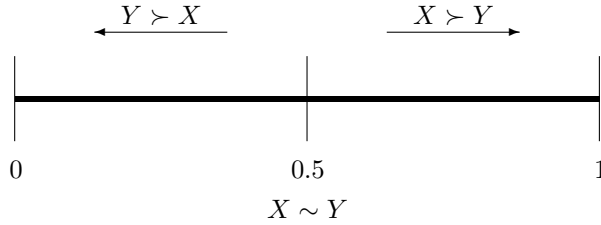


Figure 2.3: Interpretation of the reciprocal relation Q .

Statistical preference is defined from the probabilistic relation Q of Equation (2.7) and it is the formal interpretation of that relation.

Definition 2.16 ([55, 57]) *Let X and Y be two random variables. It is said that:*

- X is statistically preferred to Y , and it is denoted by $X \succeq_{\text{SP}} Y$, if $Q(X, Y) \geq \frac{1}{2}$.

Also, according to Definition 2.1:

- X and Y are statistically indifferent, and it is denoted by $X \equiv_{\text{SP}} Y$, if $Q(X, Y) = \frac{1}{2}$.
- X is strictly statistically preferred to Y , and we denote it $X \succ_{\text{SP}} Y$, if $Q(X, Y) > \frac{1}{2}$.

Note that statistical preference does not allow incomparability, so $(\mathcal{D}, \succ_{\text{SP}}, \equiv_{\text{SP}})$ constitutes a preference structure without incomparable elements.

Remark 2.17 *Statistical preference is a reflexive and complete relation. However, it is neither antisymmetric nor transitive, as we shall see in Section 3.3.♦*

It is possible to give a geometrical interpretation to the concept of statistical preference. As we can see in Figure 2.4, given two continuous and independent random variables, $X \succeq_{\text{SP}} Y$ if and only if the volume enclosed under the joint density function in the half-space $\{(x, y) \mid x > y\}$ is larger than the volume enclosed in the half-space $\{(x, y) \mid x < y\}$.

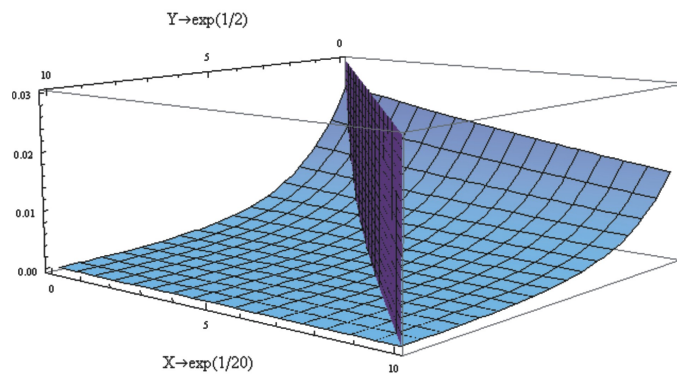


Figure 2.4: Geometrical interpretation of the statistical preference: $X \succeq_{\text{SP}} Y$.

Note that $X \succeq_{\text{SP}} Y$ means that X outperforms Y with a probability at least 0.5. Hence, statistical preference provides an order between the random variables and a preference degree. This is illustrated in the following example.

Example 2.18 *Consider two random variables X, Y such that X follows a Bernoulli distribution $\mathcal{B}(p)$ with parameter $p \in (0, 1)$ and Y follows a uniform distribution $\mathcal{U}(0, 1)$ in the interval $(0, 1)$. It is immediate that:*

$$Q(X, Y) = P(X > Y) = P(X = 1) = p.$$

Therefore, when $p \geq \frac{1}{2}$, X is statistically preferred to Y with degree of preference p , and the greater the value of p , the most preferred X is to Y .♦

One important remark is that statistical preference for degenerate random variables is equivalent to the order between real numbers, and in that case the preference degree is always 0, 1 or $\frac{1}{2}$.

Remark 2.19 Consider two random variables X and Y . The former takes the value c_X with probability 1 and the second takes the value c_Y with probability 1. Assume that $c_X > c_Y$:

$$P(X > Y) = P(X = c_X) = 1 \Rightarrow Q(X, Y) = 1 \text{ and } X \succeq_{\text{SP}} Y.$$

On the other hand, if $c_X = c_Y$, it holds that:

$$P(X = Y) = P(X = c_X, Y = c_Y) = 1 \Rightarrow Q(X, Y) = \frac{1}{2} \text{ and } X \equiv_{\text{SP}} Y.$$

Then, it holds that:

$$X \succ_{\text{SP}} Y \Leftrightarrow c_X > c_Y \text{ and } X \equiv_{\text{SP}} Y \Leftrightarrow c_X = c_Y. \blacklozenge$$

A first, but also trivial result about statistical preference is the following.

Lemma 2.20 Given two random variables X and Y , it holds that:

$$\begin{aligned} X \succeq_{\text{SP}} Y &\Leftrightarrow Q(X, Y) \geq Q(Y, X) \Leftrightarrow P(X \geq Y) \geq P(Y \geq X) \\ &\Leftrightarrow P(X > Y) \geq P(Y > X). \end{aligned}$$

Proof: By definition, $X \succeq_{\text{SP}} Y$ if and only if $Q(X, Y) \geq \frac{1}{2}$. Since Q is a probabilistic relation, $Q(X, Y) + Q(Y, X) = 1$. Then:

$$Q(X, Y) \geq \frac{1}{2} \Leftrightarrow Q(X, Y) \geq \frac{1}{2}(Q(X, Y) + Q(Y, X)) \Leftrightarrow Q(X, Y) \geq Q(Y, X).$$

Let us now prove the remaining equivalences.

$$\begin{aligned} X \succeq_{\text{SP}} Y &\Leftrightarrow Q(X, Y) \geq Q(Y, X) \\ &\Leftrightarrow P(X > Y) + \frac{1}{2}P(X = Y) \geq P(Y > X) + \frac{1}{2}P(X = Y) \\ &\Leftrightarrow P(X > Y) \geq P(Y > X). \end{aligned}$$

Moreover:

$$\begin{aligned} X \succeq_{\text{SP}} Y &\Leftrightarrow P(X > Y) \geq P(Y > X) \\ &\Leftrightarrow P(X > Y) + P(X = Y) \geq P(Y > X) + P(X = Y) \\ &\Leftrightarrow P(X \geq Y) \geq P(Y \geq X). \end{aligned} \blacksquare$$

Similar equivalences can be proved for the strict statistical preference:

$$\begin{aligned} X \succ_{\text{SP}} Y &\Leftrightarrow Q(X, Y) > Q(Y, X) \Leftrightarrow P(X \geq Y) > P(Y \geq X) \\ &\Leftrightarrow P(X > Y) > P(Y > X). \end{aligned}$$

Remark 2.21 *One context where statistical preference appears naturally is that of decision making with qualitative random variables. Dubois et al. showed in [67] that given two random variables $X, Y : \Omega \rightarrow \Omega'$, where $(\Omega', \succeq_{\Omega'})$ is an ordered qualitative scale, then, given a number of rationality axioms over our decision rule, the choice between X and Y must be made by means of the likely dominance rule, which says that X is preferred to Y if and only if $[X \succeq_{\Omega'} Y] \succsim [Y \succeq_{\Omega'} X]$, where:*

$$\begin{aligned} [X \succeq_{\Omega'} Y] &= \{\omega \in \Omega : X(\omega) \succeq_{\Omega'} Y(\omega)\} \text{ and} \\ [Y \succeq_{\Omega'} X] &= \{\omega \in \Omega : Y(\omega) \succeq_{\Omega'} X(\omega)\}, \end{aligned}$$

where \succsim is a binary relation on subsets of Ω . One of the most interesting cases is that where \succsim is determined by a probability measure P , so $A \succsim B \Leftrightarrow P(A) \geq P(B)$. Then, using Lemma 2.20, X is preferred to Y if and only if $X \succeq_{\text{SP}} Y$.

We conclude that, according to the axioms considered in [67], statistical preference is the optimal method for comparing qualitative random variables defined on a probability space. \blacklozenge

Remark 2.22 *A related notion to statistical preference is that of probability dominance considered in [210]: X is said to dominate Y with probability $\beta \geq 0.5$, and it is denoted by $X\beta Y$, if $P(X > Y) \geq \beta$. This definition has an important drawback with respect to statistical preference, which is that incomparability is possible for every $\beta \geq 0.5$. For instance, this is the case of random variables X and Y satisfying $P(X = Y) > 0.5$.*

In [2], X is called preferred to Y in the precedence order when $P(X \geq Y) \geq \frac{1}{2}$. The drawback of this notion is that indifference is possible although $P(X > Y) > P(Y > X)$, for instance when $P(X = Y) \geq \frac{1}{2}$.

From Lemma 2.20 we know that $X \succeq_{\text{SP}} Y$ if and only if $P(X > Y) \geq P(Y > X)$. When this inequality holds some authors say that X is preferred to Y in the precedence order (see [25, 26, 112]). Hence, this provides an equivalent formulation of statistical preference. We have preferred to use the latter because it provides degrees of preference between the alternatives by means of the probabilistic relation Q . Note that other authors consider a difference definition of precedence order ([2, 25, 26, 112, 210]) which is not equivalent in general, as we have seen in the previous remark. \blacklozenge

A probabilistic or reciprocal relation can also be seen as a fuzzy relation. For this reason, statistical preference can be interpreted as a defuzzification of the relation Q :

$$X \succeq_{\text{SP}} Y \Leftrightarrow (X, Y) \in Q_{\frac{1}{2}},$$

where $Q_{\frac{1}{2}}$ denotes the $\frac{1}{2}$ -cut of Q :

$$Q_{\frac{1}{2}} = \left\{ (X, Y) \in \mathcal{D} \times \mathcal{D} : Q(X, Y) \geq \frac{1}{2} \right\}. \blacklozenge$$

Another connection with fuzzy set theory can be made if we consider that the information contained in the probabilistic relation can be also presented by means of a fuzzy relation. This was initially proposed in [16, 57] and latter analyzed in detail in [122]; recently, a generalization has been presented in [163]. There, from any probabilistic relation Q defined on a set \mathcal{D} , $h(Q)$, with $h : [0, 1] \rightarrow [0, 1]$, is a fuzzy weak preference relation if and only if $h(\frac{1}{2}) = 1$.

The previous result was proven for any probabilistic relation Q , but when we are comparing random variables by means of the relation Q defined on Equation (2.7), $h(Q)$ is an order-preserving fuzzy weak preference relation if and only if $h(0) = 0$, $h(\frac{1}{2}) = 1$ and h is increasing in $[0, 1]$.

The initial h proposed in [57] was $h(x) = \min(1, 2x)$ but, of course, an infinite family of functions may be considered. As an example, we will obtain the expression of the weak preference relation R in that initial case:

$$R(X, Y) = \begin{cases} 1 & \text{if } P(X > Y) \geq P(Y > X), \\ 1 + P(X > Y) - P(Y > X) & \text{otherwise.} \end{cases}$$

Example 2.23 *Let us consider the random variable X uniformly distributed in the interval $(4, 6)$, and let Y_1 , Y_2 , Y_3 and Y_4 be the uniformly distributed random variables in the intervals $(7, 9)$, $(5, 7)$, $(3, 5)$ and $(0, 2)$, respectively. If we assume them to be independent, it holds that:*

$$\begin{aligned} Q(X, Y_1) = 0 &\Rightarrow R(X, Y_1) = 0. \\ Q(X, Y_2) = \frac{1}{8} &\Rightarrow R(X, Y_2) = \frac{1}{4}. \\ Q(X, Y_3) = \frac{7}{8} &\Rightarrow R(X, Y_3) = 1. \\ Q(X, Y_4) = 1 &\Rightarrow R(X, Y_4) = 1. \end{aligned}$$

We can notice the different scales used by Q and R . ♦

Thus, we conclude that R can be seen as a “greater than or equal to” relation, but the meaning of Q is totally different. In fact, the interpretation of the value of the fuzzy relation R is: the closer the value to 0, the weaker the preference of X over Y .

We have already mentioned some advantages of statistical preference over stochastic dominance: on the one hand, statistical preference allows the possibility of establishing preference degrees between the alternatives; on the other hand statistical preference determines a total relationship between the random variables, while we can find pairs of random variables which are incomparable under the n -th degree stochastic dominance. Another advantage is that it takes into account the possible dependence between the random variables since it is based on the joint distribution, while stochastic dominance only uses marginal distributions.

In this sense, recall that given n independent real-valued random variables X_1, \dots, X_n , with cumulative distribution functions F_{X_1}, \dots, F_{X_n} , respectively, the joint cumulative distribution function, denoted by F , is the product of the marginals:

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n),$$

for any $x_1, \dots, x_n \in \mathbb{R}$. In general, the joint cumulative distribution function can be expressed by:

$$F(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n))$$

for any $x_1, \dots, x_n \in \mathbb{R}$, where C is a function called copula.

Definition 2.24 ([166]) *A n -dimensional copula is a function $C : [0, 1]^n \rightarrow [0, 1]$ satisfying the following properties:*

- For every $(x_1, \dots, x_n) \in [0, 1]^n$, $C(x_1, \dots, x_n) = 0$ if $x_i = 0$ for some $i \in \{1, \dots, n\}$.
- For every $(x_1, \dots, x_n) \in [0, 1]^n$, $C(x_1, \dots, x_n) = x_i$ if $x_j = 1$ for every $j \neq i$.
- For every $\vec{x} = (x_1, \dots, x_n), \vec{y} = (y_1, \dots, y_n) \in [0, 1]^n$:

$$V_C([\vec{x}, \vec{y}]) \geq 0,$$

where:

$$V_C([\vec{x}, \vec{y}]) = \sum_{i=1}^n \sum_{c_i \in \{a_i, b_i\}} \text{sgn}(c_1, \dots, c_n) C(c_1, \dots, c_n),$$

where the function sgn is defined by:

$$\text{sgn}(c_1, \dots, c_n) = \begin{cases} 1 & \text{if } c_i = a_i \text{ for an even number of } i \text{'s.} \\ -1 & \text{if } c_i = a_i \text{ for an odd number of } i \text{'s.} \end{cases}$$

In particular, a 2-dimensional copula (a copula, for short) is a function $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ satisfying $C(x, 0) = C(0, x) = 0$ and $C(x, 1) = C(1, x) = x$ for every $x \in [0, 1]$ and

$$C(x_1, y_1) + C(x_2, y_2) \geq C(x_1, y_2) + C(x_2, y_1)$$

for every $(x_1, x_2, y_1, y_2) \in [0, 1]^4$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$.

The most important examples of copulas are the following:

- The product copula π : $\pi(x_1, \dots, x_n) = \prod_{i=1}^n x_i$.
- The minimum operator M : $M(x_1, \dots, x_n) = \min\{x_1, \dots, x_n\}$.
- The Łukasiewicz operator W , for $n = 2$: $W(x_1, x_2) = \max\{0, x_1 + x_2 - 1\}$.

Since the Łukasiewicz operator is associative, it can only be defined as a n -ary operator: $W(x_1, \dots, x_n) = \max\{0, x_1 + \dots + x_n - (n-1)\}$. However, it is a copula only for $n = 2$. One important and well-known result concerning copula is that every n -dimensional copula is bounded by the Łukasiewicz and the minimum operator:

$$W(x_1, \dots, x_n) \leq C(x_1, \dots, x_n) \leq M(x_1, \dots, x_n) \text{ for every } (x_1, \dots, x_n) \in [0, 1]^n. \quad (2.8)$$

This inequality is known as the Fréchet-Hoeffding inequality. For this reason, the Łukasiewicz and the minimum operators are also called the *lower and upper Fréchet-Hoeffding bounds* ([79]).

Recall that, although W is not a copula for $n > 2$, it can be approximated by a copula on each point:

Proposition 2.25 ([62, 166]) *For any $(x_1, \dots, x_n) \in [0, 1]^n$ there is a n -dimensional copula C such that $C(x_1, \dots, x_n) = W(x_1, \dots, x_n)$.*

In particular, when $n = 2$, W is a copula and the previous result becomes trivial.

A particular type of copulas are the Archimedean copulas.

Definition 2.26 ([166]) *A n -dimensional copula C is Archimedean if there exists a function $\varphi : [0, 1] \rightarrow [0, \infty]$, called generator of C , strictly decreasing, satisfying that $-\varphi$ is n -monotone, $\varphi(1) = 0$ and:*

$$C(x_1, \dots, x_n) = \varphi^{-1}(\varphi(x_1) + \dots + \varphi(x_n)), \quad (2.9)$$

for every $(x_1, \dots, x_n) \in [0, 1]^n$, where φ^{-1} denotes the pseudo-inverse of φ , and it is defined by:

$$\varphi^{-1}(t) = \begin{cases} \varphi^{-1}(t) & \text{if } 0 \leq t \leq \varphi(0). \\ 0 & \text{if } \varphi(0) < t \leq \infty. \end{cases}$$

The main Archimedean copulas are the product, whose generator is $\varphi_\pi(t) = -\log t$, and the Łukasiewicz operator for $n = 2$, whose generator is $\varphi_W(t) = 1 - t$. The most important non-Archimedean copula is the minimum operator.

Archimedean copulas can also be divided into two groups: strict and nilpotent Archimedean copulas. An Archimedean copula is called *strict* if its generator, φ , satisfies $\varphi(0) = \infty$. In such case, the pseudo inverse becomes the inverse, and therefore Equation (2.9) becomes:

$$C(x_1, \dots, x_n) = \varphi^{-1}(\varphi(x_1) + \dots + \varphi(x_n)). \quad (2.10)$$

An Archimedean copula is *nilpotent* if $\varphi(0) < \infty$. The most important examples of strict and nilpotent copulas are the product and the Łukasiewicz operator, respectively.

One of the most important traits of copulas is the famous Sklar's theorem.

Theorem 2.27 ([189]) *Let X_1, \dots, X_n be n random variables, and let F_{X_1}, \dots, F_{X_n} denote their respective cumulative distribution functions. If F denotes the joint cumulative distribution function, then there exists a copula C such that*

$$F(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) \text{ for every } (x_1, \dots, x_n) \in \mathbb{R}^n.$$

When the copula is Archimedean, last expression becomes:

$$F(x_1, \dots, x_n) = \varphi^{-1}(\varphi(F_{X_1}(x_1)) + \dots + \varphi(F_{X_n}(x_n))).$$

Obviously, a pair of random variables is coupled by the product if and only if they are independent. Moreover, random variables coupled by the minimum operator (respectively, by the Łukasiewicz operator) are called *comonotonic* (respectively, *countermonotonic*). These two cases are very important in the theory of copulas, and for this reason we will study in detail the properties of statistical preference and stochastic dominance for them. In fact, from the Fréchet-Hoeffding bounds of Equation (2.8), an interpretation of comonotonic and countermonotonic random variables can be given. In order to see this, recall that a subset S of $\overline{\mathbb{R}}^2$ is increasing if and only if for each $(x, y) \in \mathbb{R}^2$ either:

1. for all (u, v) in S , $u \leq x$ implies $v \leq y$; or
2. for all (u, v) in S , $v \leq y$ implies $u \leq x$.

Similarly, a subset S of $\overline{\mathbb{R}}^2$ is decreasing if and only if for each $(x, y) \in \mathbb{R}^2$ either:

1. for all (u, v) in S , $u \leq x$ implies $v \geq y$; or
2. for all (u, v) in S , $v \leq y$ implies $u \geq x$.

Using this notation, the following result is presented in [166, Theorem 2.5.4] and proved in [124].

Proposition 2.28 *Let X and Y be two real-valued random variables. X and Y are comonotonic if and only if the support of the joint distribution function is a increasing subset of $\overline{\mathbb{R}}^2$, and X and Y are countermonotonic if and only if the support of the joint distribution function is a decreasing subset of $\overline{\mathbb{R}}^2$.*

When X and Y are continuous, we say that Y is almost surely an increasing function of X if and only if X and Y are comonotonic, and Y is almost surely a decreasing function of X if and only if they are countermonotonic.

2.2 Imprecise probabilities

Next, we discuss briefly *imprecise probability models*. This is the generic term used to refer to all mathematical models that serve as an alternative and a generalization of probability models to situations where our knowledge is vague or scarce. It includes possibility measures ([217]), Choquet capacities ([39]), belief functions ([187]) or coherent lower previsions ([205]), among other models.

2.2.1 Coherent lower previsions

We begin by introducing the main concepts of the theory of coherent lower previsions. Consider a possibility space Ω . A *gamble* is a real-valued functional defined on Ω . We shall denote by $\mathcal{L}(\Omega)$ the set of all gambles on Ω , while $\mathcal{L}^+(\Omega)$ denotes the set of positive gambles on Ω . Given a subset A of Ω , the indicator function of A is the gamble that takes the value 1 on the elements of A and 0 elsewhere. We shall denote this gamble by I_A , or by A when no confusion is possible.

A *lower prevision* is a functional \underline{P} defined on a set of gambles $\mathcal{K} \subseteq \mathcal{L}(\Omega)$. Given a gamble f , $\underline{P}(f)$ is understood to represent a subject's supremum acceptable buying price for f , in the sense that for any $\varepsilon > 0$ the transaction $f - \underline{P}(f) + \varepsilon$ is acceptable to him.

Using this interpretation, we can derive the notion of coherence.

Definition 2.29 ([205, Section 2.5]) *Consider the lower prevision $\underline{P} : \mathcal{K} \rightarrow \mathbb{R}$, where $\mathcal{K} \subseteq \mathcal{P}(\Omega)$. It avoids sure loss if for any natural number n and any $f_1, \dots, f_n \in \mathcal{K}$ it holds that:*

$$\sup_{\omega \in \Omega} \left[\sum_{k=1}^n [f_k(\omega) - \underline{P}(f_k)] \right] \geq 0.$$

Also, \underline{P} is coherent if for any natural numbers n and m and $f_0, f_1, \dots, f_n \in \mathcal{K}$, it holds that:

$$\sup_{\omega \in \Omega} \left[\sum_{i=1}^n [f_i(\omega) - \underline{P}(f_i)] - m[f_0(\omega) - \underline{P}(f_0)] \right] \geq 0.$$

The interpretation of this notion is that the acceptable buying prices encompassed by $\{\underline{P}(f) : f \in \mathcal{L}(\Omega)\}$ are consistent with each other, in the sense defined in [205, Section 2.5]. From any lower prevision \underline{P} it is possible to define a set of probabilities, also called credal set, by:

$$\mathcal{M}(\underline{P}) = \{P \text{ finitely additive probabilities} : P \geq \underline{P}\}.$$

The following result relates coherence and avoiding sure loss to the credal set $\mathcal{M}(\underline{P})$. It is usually called the *Envelope Theorem*.

Theorem 2.30 ([205, Section 3.3.3]) *Let \underline{P} be a lower probability defined on a set of gambles \mathcal{K} , and let $\mathcal{M}(\underline{P})$ denote its associated credal set. Then:*

$$\underline{P} \text{ avoids sure loss} \Leftrightarrow \mathcal{M}(\underline{P}) \neq \emptyset$$

and

$$\underline{P} \text{ is coherent} \Leftrightarrow \underline{P}(f) = \inf_{P \in \mathcal{M}(\underline{P})} P(f).$$

By conjugacy, an operator \overline{P} defined on a set of gambles \mathcal{K} is called *upper prevision*. For any $f \in \mathcal{K}$, $\overline{P}(f)$ is understood to represent the subject's infimum acceptable selling price for f , in the sense that for any $\varepsilon > 0$ the transaction $\overline{P}(f) + \varepsilon - f$ is acceptable to him. An upper prevision avoids sure loss (respectively, is coherent) if and only if $\overline{P}(f) = -\underline{P}(-f)$, where \underline{P} is a lower prevision that avoids sure loss (respectively, that is coherent).

When the domain \mathcal{K} of the lower and upper previsions is formed by subsets of Ω , \underline{P} and \overline{P} are called *lower and upper probabilities*, respectively.

Next proposition shows several properties of coherent lower and upper probabilities.

Proposition 2.31 ([205, Section 2.4.7]) *Let \underline{P} be a lower probability and let \overline{P} denote its conjugate upper probability. The following statements hold for any $A, B \subseteq \Omega$:*

$$A \cap B = \emptyset \Rightarrow \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B). \quad (2.11)$$

$$A \cap B = \emptyset \Rightarrow \overline{P}(A \cup B) \geq \underline{P}(A) + \overline{P}(B). \quad (2.12)$$

$$\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) + \overline{P}(A \cap B). \quad (2.13)$$

$$\overline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B). \quad (2.14)$$

$$\overline{P}(A \cup B) + \overline{P}(A \cap B) \geq \overline{P}(A) + \underline{P}(B). \quad (2.15)$$

Given a coherent lower prevision \underline{P} with domain \mathcal{K} , we may be interested in extending \underline{P} to a more general domain $\mathcal{K}' \supseteq \mathcal{K}$. This can be made by means of the natural extension.

Definition 2.32 ([205, Section 3.1]) *Let \underline{P} be a coherent lower prevision on \mathcal{K} , and consider $\mathcal{K}' \supseteq \mathcal{K}$. Then, for any $f \in \mathcal{K}'$, the natural extension of \underline{P} is defined by:*

$$\underline{E}(f) = \inf_{P \in \mathcal{M}(\underline{P})} P(f).$$

The natural extension is the least committal, that is the most imprecise, coherent extension of \underline{P} .

One instance where coherent lower previsions appear is when dealing with p-boxes.

Definition 2.33 ([75]) A probability box, or *p-box* for short, $(\underline{F}, \overline{F})$ is the set of cumulative distribution functions bounded between two finitely additive distribution functions \underline{F} and \overline{F} such that $\underline{F} \leq \overline{F}$. We shall refer to \underline{F} as the lower distribution function and to \overline{F} as the upper distribution function of the *p-box*.

Note that $\underline{F}, \overline{F}$ need not be cumulative distribution functions, and as such they need not belong to the set $(\underline{F}, \overline{F})$; they are only required to be finitely additive distribution functions. In particular, if we consider a set \mathcal{F} of distribution functions, its associated lower and upper distribution functions are given by

$$\underline{F}(x) := \inf_{F \in \mathcal{F}} F(x), \quad \overline{F}(x) := \sup_{F \in \mathcal{F}} F(x). \quad (2.16)$$

Proposition 2.34 Given a set of cumulative distribution functions \mathcal{F} , its lower bound \underline{F} is also a cumulative distribution function, while \overline{F} is a finitely additive cumulative distribution function.

P-boxes have been connected to info-gap theory ([76]), random sets ([103, 172]), and possibility measures ([17, 51, 198]).

Given a *p-box* $(\underline{F}, \overline{F})$ on Ω , it induces a lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ on the set

$$\mathcal{K} = \{A_x, X_x^c : x \in \Omega\},$$

where $A_x = \{x' \in \Omega : x' \leq x\}$, by:

$$\underline{P}_{(\underline{F}, \overline{F})}(A_x) = \underline{F}(x) \text{ and } \underline{P}_{(\underline{F}, \overline{F})}(A_x^c) = 1 - \overline{F}(x). \quad (2.17)$$

If $\underline{F} = \overline{F} = F$, $\underline{P}_{(\underline{F}, \overline{F})}$ is usually denoted by P_F . The following result is stated in [209] and proved in [198, 201].

Theorem 2.35 ([198, Section 3],[201, Theorem 3.59]) Consider two maps \underline{F} and \overline{F} from Ω to $[0, 1]$ and let $\underline{P}_{(\underline{F}, \overline{F})} : \mathcal{K} \rightarrow [0, 1]$ be the lower probability they induce by means of Equation (2.17). The following statements are equivalent:

- $\underline{P}_{(\underline{F}, \overline{F})}$ is a coherent lower probability.
- $\underline{F}, \overline{F}$ are distribution functions and $\underline{F} \leq \overline{F}$.
- $P_{\underline{F}}$ and $P_{\overline{F}}$ are coherent and $\underline{F} \leq \overline{F}$.

In particular, if $\underline{F} = \overline{F} = F$, then P_F is coherent if and only if F is a distribution function.

A particular case appears when defining coherent lower previsions in product spaces $\Omega_1 \times \Omega_2$. If \underline{P} is a coherent lower prevision taking values on $\mathcal{L}(\Omega_1 \times \Omega_2)$, we can consider its marginals \underline{P}_1 or \underline{P}_2 as coherent lower previsions on $\mathcal{L}(\Omega_1)$ or $\mathcal{L}(\Omega_2)$, respectively, defined by:

$$\underline{P}_1(f) = \underline{P}(f) \text{ and } \underline{P}_2(f) = \underline{P}(f)$$

for any gamble f on $\Omega_1 \times \Omega_2$. They will arise when trying to define coherent lower previsions from bivariate p-boxes.

In this work, we shall use imprecise probability models because we shall be interested in the comparison of sets of alternatives, each with its associated probability distribution; we obtain thus a set \mathcal{P} of probability measures. This set can be summarized by means of its *lower* and *upper* envelopes, which are given by:

$$\underline{P}(A) := \inf_{P \in \mathcal{P}} P(A), \quad \overline{P}(A) := \sup_{P \in \mathcal{P}} P(A), \quad (2.18)$$

and which are *coherent* lower and upper probabilities.

2.2.2 Conditional lower previsions

Consider two random variables X and Y taking values in two spaces Ω_1 and Ω_2 and let \underline{P} be a coherent lower prevision taking values on $\mathcal{L}(\Omega_1 \times \Omega_2)$. We define a conditional lower prevision $\underline{P}(\cdot | Y)$ as a function with two arguments. For any $y \in \Omega_2$, $\underline{P}(\cdot | y)$ is real functional on the set $\mathcal{L}(\Omega_1 \times \Omega_2)$, while for any gamble f on $\Omega_1 \times \Omega_2$, $\underline{P}(f | y)$ is the lower prevision of f , conditional on $\Omega_2 = y$. $\underline{P}(f | Y)$ is then the gamble on Ω_1 that assumes the value $\underline{P}(f | y)$ in y . Similar considerations can be made for $\underline{P}(\cdot | X)$.

Definition 2.36 *The conditional lower prevision $\underline{P}(\cdot | Y)$ is called separately coherent if for all $y \in \Omega_2$, $\lambda \geq 0$ and $f, g \in \mathcal{L}(\Omega_1 \times \Omega_2)$ it satisfies the following conditions:*

- SC1 $\underline{P}(f | y) \geq \inf_{x \in \Omega} f(x, y)$.
- SC2 $\underline{P}(\lambda f | y) = \lambda \underline{P}(f | y)$.
- SC3 $\underline{P}(f + g | y) \geq \underline{P}(f | y) + \underline{P}(g | y)$.

It is known that from separate coherence the following properties hold (see [205, Theorems 6.2.4 and 6.2.6]):

$$\underline{P}(g | y) = \underline{P}(g(\cdot, y) | y) \text{ and } \underline{P}(fg | Y) = f \underline{P}(g | Y),$$

for all $y \in \Omega_2$, all positive gambles f on Ω_2 and all gambles g on $\Omega_1 \times \Omega_2$.

We now investigate separate coherence and coherence together. For any gamble f on $\mathcal{L}(\Omega_1 \times \Omega_2)$, we define:

$$G(f | y) = I_{\{y\}}[f - \underline{P}(f | Y)] = I_{\{y\}}[f(\cdot, y) - \underline{P}(f(\cdot, y) | y)]$$

and

$$G(f | Y) = f - \underline{P}(f | Y) = f - \underline{P}(f | Y) = \sum_{y \in \Omega'} I_{\{y\}} [f(\cdot, y) - \underline{P}(f(\cdot, y) | y)].$$

Definition 2.37 Let $\underline{P}(\cdot | Y)$ and $\underline{P}(\cdot | X)$ be two separately coherent conditional lower previsions. They are called weakly coherent if and only if for all $f_1, f_2 \in \mathcal{L}(\Omega_1 \times \Omega_2)$, all $x \in \Omega_1, y \in \Omega_2$ and $g \in \mathcal{L}(\Omega_1 \times \Omega_2)$, there are some

$$\begin{aligned} B_1 &\subseteq \text{supp}_{\Omega_1}(f_2) \cup \text{supp}_{\Omega_2}(f_1) \cup (\{x\} \times \Omega_2) \\ B_2 &\subseteq \text{supp}_{\Omega_1}(f_2) \cup \text{supp}_{\Omega_2}(f_1) \cup (\Omega_1 \times \{y\}) \end{aligned}$$

such that:

$$\sup_{z \in B_1} [G(f_1 | Y) + G(f_2 | X) - G(g | x)](z) \geq 0$$

and

$$\sup_{z \in B_2} [G(f_1 | Y) + G(f_2 | X) - G(g | y)](z) \geq 0,$$

where

$$\text{supp}_{\Omega_1}(f) = \{\{x\} \times \Omega_2, x \in \Omega_1 \mid f(x, \cdot) \neq 0\}$$

and

$$\text{supp}_{\Omega_2}(f) = \{\Omega_1 \times \{y\}, y \in \Omega_2 \mid f(\cdot, y) \neq 0\}.$$

We say that $\underline{P}(\cdot | Y)$ and $\underline{P}(\cdot | X)$ are coherent if for all $f_1, f_2 \in \mathcal{L}(\Omega_1 \times \Omega_2)$, all $x \in \Omega_1, y \in \Omega_2$ and all $g \in \mathcal{L}(\Omega_1 \times \Omega_2)$ it holds that:

$$\begin{aligned} \sup_{z \in \Omega_1 \times \Omega_2} [G(f_1 | Y) + G(f_2 | X) - G(g | x)](z) &\geq 0. \\ \sup_{z \in \Omega_1 \times \Omega_2} [G(f_1 | Y) + G(f_2 | X) - G(g | y)](z) &\geq 0. \end{aligned}$$

Several results can be found in the literature relating coherence and weak coherence.

Theorem 2.38 ([137, Theorem 1]) Let $\underline{P}(\cdot | X)$ and $\underline{P}(\cdot | Y)$ be separately coherent conditional lower previsions. They are weakly coherent if and only if there is some coherent lower prevision \underline{P} on $\mathcal{L}(\Omega_1 \times \Omega_2)$ such that

$$\begin{aligned} \underline{P}(G(f | X)) \geq 0 \text{ and } \underline{P}(G(f | x)) = 0 \text{ for any } f \in \mathcal{L}(\Omega_2), x \in \Omega_2, \\ \underline{P}(G(g | Y)) \geq 0 \text{ and } \underline{P}(G(g | y)) = 0 \text{ for any } g \in \mathcal{L}(\Omega_1), y \in \Omega_1. \end{aligned}$$

The following result is known as the Reduction Theorem.

Theorem 2.39 ([205, Theorem 7.1.5]) Let $\underline{P}(\cdot | X)$ and $\underline{P}(\cdot | Y)$ be separately coherent conditional lower previsions defined on $\mathcal{L}(\Omega_1 \times \Omega_2)$, and let \underline{P} be a coherent lower prevision on $\mathcal{L}(\Omega_1 \times \Omega_2)$. Then $\underline{P}, \underline{P}(\cdot | X)$ and $\underline{P}(\cdot | Y)$ are coherent if and only if the following two conditions holds:

1. $\underline{P}, \underline{P}(\cdot | X)$ and $\underline{P}(\cdot | Y)$ are weakly coherent.
2. $\underline{P}(\cdot | X)$ and $\underline{P}(\cdot | Y)$ are coherent.

2.2.3 Non-additive measures

One important example of coherent lower previsions are the n -monotone ones, which were first introduced by Choquet in [39].

Definition 2.40 ([39]) *A coherent lower prevision \underline{P} on $\mathcal{L}(\Omega)$ is called n -monotone if and only if:*

$$\underline{P}\left(\bigvee_{i=1}^p f_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, p\}} (-1)^{|I|+1} \underline{P}\left(\bigwedge_{i \in I} f_i\right)$$

for all $2 \leq p \leq n$ and all f_1, \dots, f_p in $\mathcal{L}(\Omega)$, where \vee denotes the point-wise maximum and \wedge the point-wise minimum.

In particular, a coherent lower probability $\underline{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ is n -monotone when

$$\underline{P}\left(\bigcup_{i=1}^p A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, p\}} (-1)^{|I|+1} \underline{P}\left(\bigcap_{i \in I} A_i\right)$$

for all $2 \leq p \leq n$ and all subsets A_1, \dots, A_p of Ω .

A coherent lower prevision on $\mathcal{L}(\Omega)$, that is n -monotone for all $n \in \mathbb{N}$, is called *completely monotone*, and its restriction to events is a *belief function*. The restriction to events of the conjugate upper prevision is called *plausibility function*. Belief and plausibility functions are usually denoted by *bel* and *pl*.

Another type of non-additive measure are possibility measures.

Definition 2.41 ([70]) *A possibility measure on $[0, 1]$ is a supremum preserving set function $\Pi : \mathcal{P}([0, 1]) \rightarrow [0, 1]$. It is characterised by its restriction to events π , which is called its possibility distribution. The conjugate function N of a possibility measure is called a necessity measure:*

$$N(A) = 1 - \Pi(A^c).$$

Because of their computational simplicity, possibility measures are widely applied in many fields, including data analysis ([196]), diagnosis ([33]), case-based reasoning ([91]) and psychology ([177]).

Let us see how to apply our extension stochastic dominance to the comparison of possibility measures; another approach to preference modeling with possibility measures is discussed in [19, 115].

The connection between possibility measures and p-boxes was already explored in [199], and it was proven that almost any possibility measure can be seen as the natural

extension of a corresponding p-box. However, the definition of this p-box implies defining some particular order on our referential space, which could be different to the one we already have there (for instance if the possibility measure is defined on $[0,1]$ it may seem counterintuitive to consider anything different from the natural order), and moreover two different possibility measures may produce two different orders on the same space, making it impossible to compare them.

Instead, we shall consider a possibility measure Π on $\Omega = [0,1]$, its associated set of probability measures:

$$\mathcal{M}(\Pi) := \{P \text{ probability} : P(A) \leq \Pi(A) \forall A\}, \quad (2.19)$$

and the corresponding set of distribution functions \mathcal{F} . Let $(\underline{F}, \overline{F})$ be its associated p-box.

Since any possibility measure on $[0,1]$ can be obtained as the upper probability of a random set ([84]), and moreover in that case ([131]) the upper probability of the random set is the maximum of the probability distributions of the measurable selections, we deduce that the p-box associated to \mathcal{F} is determined by the following lower and upper distribution functions:

$$\begin{aligned} \overline{F}(x) &= \sup_{P \leq \Pi} P([0, x]) = \Pi([0, x]) = \sup_{y \leq x} \pi(y) \\ \underline{F}(x) &= \inf_{P \leq \Pi} P([0, x]) = 1 - \Pi((x, 1]) = 1 - \sup_{y > x} \pi(y). \end{aligned} \quad (2.20)$$

Note however, that these lower and upper distribution functions need not belong to \mathcal{F} : if for instance we consider the possibility measure associated to the possibility distribution $\pi = I_{(0.5, 1]}$, we obtain $\overline{F} = \pi$, which is not right-continuous, and consequently cannot belong to the set \mathcal{F} of distribution functions associated to $\mathcal{M}(\Pi)$.

Another interesting type of non-additivity measures, that includes possibility measures as a particular case are clouds. Following Neumaier ([168]), a *cloud* is a pair of functions $[\delta, \pi]$, where $\pi, \delta : [0, 1] \rightarrow [0, 1]$ satisfy:

- $\delta \leq \pi$.
- There exists $x \in [0, 1]$ such that $\pi(x) = 0$.
- There exists $y \in [0, 1]$ such that $\delta(y) = 1$.

δ and π are called the *lower* and *upper distributions* of the cloud, respectively.

Any cloud $[\delta, \pi]$ has an associated set of probabilities $\mathcal{P}_{[\delta, \pi]}$, that is the set of probabilities P satisfying:

$$P(\{x \in [0, 1] \mid \delta(x) \geq \alpha\}) \leq 1 - \alpha \leq P(\{x \in [0, 1] \mid \pi(x) > \alpha\}).$$

Since both π and $1-\delta$ are possibility distributions we can consider their associated credal sets \mathcal{P}_π and $\mathcal{P}_{1-\delta}$, given by

$$\mathcal{P}_\pi := \{P \text{ probability} : P(A) \leq \Pi(A) \forall A \in \beta_{[0,1]}\},$$

where Π denotes the possibility measure associated to the possibility distribution π , and similarly for $\mathcal{P}_{1-\delta}$. From [65], it holds that $\mathcal{P}_{[\delta,\pi]} = \mathcal{P}_{1-\delta} \cap \mathcal{P}_\pi$.

2.2.4 Random sets

One context where completely monotone lower previsions arise naturally is that of measurable multi-valued mappings, or random sets ([59, 96]).

Definition 2.42 *Let (Ω, \mathcal{A}, P) be a probability space, (Ω', \mathcal{A}') a measurable space, and $\Gamma : \Omega \rightarrow \mathcal{P}(\Omega')$ a non-empty multi-valued mapping. It is called random set when*

$$\Gamma^*(A) = \{\omega \in \Omega : \Gamma(\omega) \cap A \neq \emptyset\} \in \mathcal{A}$$

for any $A \in \mathcal{A}'$.

One instance of random sets are random intervals, that are those satisfying that $\Gamma(\omega)$ is an interval for any $\omega \in \Omega$.

If Γ models the imprecise knowledge about a random variable X , $\Gamma(\omega)$ represents that the “true” value of $X(\omega)$ belongs to $\Gamma(\omega)$. Then, all we know about X is that it is one of the *measurable selections* of Ω :

$$S(\Gamma) = \{U : \Omega \rightarrow \Omega' \text{ random variable} : U(\omega) \in \Gamma(\omega) \forall \omega \in \Omega\}. \quad (2.21)$$

This interpretation of multi-valued mappings as a model for the imprecise knowledge of a random variable is not new, and can be traced back to Kruse and Meyer ([104]). The *epistemic* interpretation contrasts with the *ontic* interpretation which is sometimes given to random sets as naturally imprecise quantities ([73]).

Random sets generate upper and lower probabilities.

Definition 2.43 ([59]) *Let (Ω, \mathcal{A}, P) be a probability space, (Ω', \mathcal{A}') a measurable space and $\Gamma : \Omega \rightarrow \mathcal{P}(\Omega')$ a random set. Then its upper and lower probabilities are the functions $P^*, P_* : \mathcal{A}' \rightarrow [0, 1]$ given by:*

$$P^*(A) = P(\{\omega : \Gamma(\omega) \cap A \neq \emptyset\}) \text{ and } P_*(A) = P(\{\omega : \emptyset \neq \Gamma(\omega) \subseteq A\}) \quad (2.22)$$

for any $A \in \mathcal{A}'$. These upper and lower probabilities are, in particular, a *plausibility* and a *belief function*, respectively. Furthermore, they define the credal set $\mathcal{M}(P_\Gamma^*)$ given by:

$$\mathcal{M}(P_\Gamma^*) = \{P \text{ probability} : P_{*\Gamma}(A) \leq P(A) \leq P_\Gamma^*(A) \forall A \in \mathcal{A}'\}. \quad (2.23)$$

The upper and lower probabilities of a random set are in particular coherent lower and upper probabilities, and constitute the lower and upper bounds of the probabilities induced by the measurable selections:

$$P_*(A) \leq P_X(A) \leq P^*(A) \text{ for every } X \in S(\Gamma). \quad (2.24)$$

Therefore, their associated cumulative distribution functions provide lower and upper bounds of the lower and upper distribution functions associated to $S(\Gamma)$. The inequalities of Equation (2.24) can be strict [130, Example 1]; however, under fairly general conditions

$$P^*(A) = \max \mathcal{P}(\Gamma)(A) \text{ and } P_*(A) = \min \mathcal{P}(\Gamma)(A) \text{ for every } A \in \mathcal{A}', \quad (2.25)$$

where $\mathcal{P}(\Gamma)(A) = \{P_X(A) : X \in S(\Gamma)\}$. In particular, if Γ takes values on the measurable space $([0, 1], \beta_{[0,1]})$, where $\beta_{[0,1]}$ denotes the Borel σ -field, Equation (2.25) holds under any of the following conditions ([130]):

- If the class $\{\Gamma(\omega) : \omega \in \Omega\}$ is countable.
- If $\Gamma(\omega)$ is closed for every $\omega \in \Omega$.
- If $\Gamma(\omega)$ is open for every $\omega \in \Omega$.

However, the two sets are not equivalent in general, and $\mathcal{M}(P_\Gamma^*)$ can only be seen as an outer approximation. There are nonetheless situations in which both sets coincide. First, let us introduce the following definition.

Definition 2.44 *Consider two functions $A, B : \Omega \rightarrow \mathbb{R}$. They are called strictly comonotone if $(A(\omega) - A(\omega')) \geq 0$ if and only if $(B(\omega) - B(\omega')) \geq 0$ for any $\omega, \omega' \in \Omega$.*

A similar but less restrictive notion is the one of comonotone functions: A and B are called comonotone if $(A(\omega) - A(\omega'))(B(\omega) - B(\omega')) \geq 0$ for any $\omega, \omega' \in \Omega$. Note that both notions are not equivalent in general. In fact, two increasing and comonotone functions A and B are strictly comonotone if and only if $A(\omega) = A(\omega')$ if and only if $B(\omega) = B(\omega')$, and two comonotone functions A and B with $A = 0 \leq B$ are strictly comonotone if and only if B is constant.

Next, we list some situations in which the sets $P^*(\Gamma)$ and $\mathcal{M}(P_\Gamma^*)$ coincide.

Proposition 2.45 ([129]) *Let (Ω, \mathcal{A}, P) be a probability space and consider the random closed interval $\Gamma := [A, B] : \Omega \rightarrow \mathcal{P}(\mathbb{R})$. Let $P(\Gamma), \mathcal{M}(P_\Gamma^*)$ denote the sets of probability measures induced by the selections and those dominated by the upper probability, respectively. Then:*

1. $P_\Gamma^*(C) = \max\{Q(C) : Q \in P(\Gamma)\} \forall C \in \beta_{\mathbb{R}}$.

2. $M(P_\Gamma^*) = \overline{\text{Conv}(P(\Gamma))}$, and if (Ω, \mathcal{A}, P) is non-atomic then $M(P_\Gamma^*) = \overline{P(\Gamma)}$.
3. When $(\Omega, \mathcal{A}, P) = ([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$, the equality $M(P_\Gamma^*) = P(\Gamma)$ holds under any of the following conditions:
 - (a) The variables $A, B : [0, 1] \rightarrow \mathbb{R}$ are increasing.
 - (b) $A = 0 \leq B$.
 - (c) A, B are strictly comonotone.

For a complete study on the conditions under which the lower and upper probabilities are attained or the conditions under which the sets $\mathcal{P}(\Gamma)$ and $\mathcal{M}(P_*)$ coincide, we refer to [125].

Theorem 2.46 ([130, Theorem 14]) *Let (Ω, \mathcal{A}, P) be a probability space. Consider the measurable space $([0, 1], \beta_{[0,1]})$ and let $\Gamma : \Omega \rightarrow \mathcal{P}([0, 1])$ be a random set. If $P^*(A) = \max \mathcal{P}(\Gamma)(A)$ for all $A \in \mathcal{A}'$, then for any bounded random variable $f : [0, 1] \rightarrow \mathbb{R}$:*

$$(C) \int f dP^* = \sup_{U \in S(\Gamma)} \int f dP_U, \quad (C) \int f dP_* = \inf_{U \in S(\Gamma)} \int f dP_U,$$

and consequently:

$$(C) \int f dP^* = \sup(A) \int (f \circ \Gamma) dP, \quad (C) \int f dP_* = \inf(A) \int (f \circ \Gamma) dP,$$

where $(C) \int f dP^*$ denotes the Choquet integral of f with respect to P^* , and $(A) \int (f \circ \Gamma) dP$ denotes the Aumann integral of $f \circ \Gamma$ with respect to P , given by:

$$(A) \int (f \circ \Gamma) dP = \left\{ \int f dP_U : U \in S(\Gamma) \right\}. \quad (2.26)$$

The upper probability induced by a random set is always completely alternating and lower continuous [169]. Under some additional conditions, it is in particular maxitive or a possibility measure:

Proposition 2.47 ([128, Corollary 5.4]) *Let (Ω, \mathcal{A}, P) be a probability space and consider the random closed interval $\Gamma : \Omega \rightarrow \mathcal{P}(\mathbb{R})$. The following are equivalent:*

- (a) P_Γ^* is a possibility measure.
- (b) P_Γ^* is maxitive.
- (c) There exists some $N \subseteq \Omega$ null such that for every $\omega_1, \omega_2 \in \Omega \setminus N$, either $\Gamma(\omega_1) \subseteq \Gamma(\omega_2)$ or $\Gamma(\omega_2) \subseteq \Gamma(\omega_1)$.

See also [50] for related results when $\Omega = [0, 1]$.

2.3 Intuitionistic fuzzy sets

Fuzzy sets were introduced by Zadeh ([214]) as a suitable model for situations where crisp sets did not convey appropriately the available information. However, there are also situations where a more general model than fuzzy sets is deemed adequate.

A fuzzy set A assigns to every point on the universe a number in $[0, 1]$ that measures the degree in which this point is compatible with the characteristic described by A . Thus, if $A(\omega)$ denotes the membership degree of ω to A , $1 - A(\omega)$ stands for the degree in which ω does not belong to A . However, two problems can arise in this situation:

1. $1 - A(\omega)$ could include at the same time both the degree of non-membership and the degree of uncertainty or indeterminacy.
2. The membership degree could not be precisely described.

Consider the following example for the former case:

Example 2.48 *Let A be the set $A = \text{“objects possessing some characteristic”}$. Thus, $A(\omega)$ stands for the degree in which ω is in accord with the given characteristic, and $1 - A(\omega)$ is the degree in which ω is not. However, ω could be partly indifferent to the characteristic. To deal with this situation, we can denote by $\mu_A(\omega) = A(\omega)$ the membership degree of ω in A , and let us define by $\nu_A(\omega)$ the degree in which ω does not belong to A . Such sets, where a membership and non-membership degree is associated with any element, are called (Atanassov) Intuitionistic Fuzzy Sets (in short, IF-sets). A good example of these situations is voting, since human voters can be grouped in three classes: vote for, vote against or abstain ([195]).♦*

In order to illustrate second scenario, consider the following example:

Example 2.49 *We are studying some element with melting temperature is m and vaporization temperature is v (obviously, $m \leq v$). For example, for water $m = 0^\circ\text{C}$ and $v = 100^\circ\text{C}$. If the element is in a liquid state, we know that its temperature is greater than m , because otherwise it would be solid, and smaller than v , because otherwise it would be in gaseous state. Then, although we cannot state the exact temperature of the element, we can say for sure that it belongs to the interval $[m, v]$.♦*

If $A(\omega)$ denotes the (non-precisely known) membership degree of ω to A , we can consider an interval $[l_A(\omega), u_A(\omega)]$ that represents that the exact membership degree of ω to A belongs to such interval. These sets, where any element has an associated interval that bounds the membership degree of the element to the set, are called Interval Valued Fuzzy Sets (IVF-sets, for short).

In this section we introduce the definition and main properties of both IF-sets and IVF-sets, and we see how the usual operations between crisp sets can be generalized into this context. In particular, we show that both kind of sets are formally equivalent although, as we have already mentioned, their philosophy is different.

Let us begin with the formal definition of an intuitionistic fuzzy set.

Definition 2.50 ([4]) *Let Ω be a universe. An intuitionistic fuzzy set is defined by:*

$$A = \{(\omega, \mu_A(\omega), \nu_A(\omega)) \mid \omega \in \Omega\},$$

where μ_A and ν_A are functions:

$$\mu_A, \nu_A : \Omega \rightarrow [0, 1]$$

satisfying $\mu_A(\omega) + \nu_A(\omega) \leq 1$. The function $\pi_A(\omega) = 1 - \mu_A(\omega) - \nu_A(\omega)$ is called the hesitation index and it expresses the lack of knowledge on the membership of ω to A .

We shall denote the set of all IF-sets on Ω by $IFSS(\Omega)$.

When A is a fuzzy set, its complementary is given by $A^c = 1 - A$. That is, the membership degree of every element to the complementary of A is one minus the membership degree to A . Then, every fuzzy set is in particular an IF-set where the hesitation index equals zero. If $FS(\Omega)$ denotes all fuzzy sets on Ω , $FS(\Omega) \subset IFSS(\Omega)$. For proper IF-sets, if μ_A and ν_A denote the membership and non-membership functions, the complementary of A is defined by:

$$A^c = \{(\omega, \nu_A(\omega), \mu_A(\omega)) \mid \omega \in \Omega\}.$$

Recall that, since the empty set \emptyset is the set with no elements, it can be also seen as an IF-set given by:

$$\emptyset = \{(\omega, 0, 1) \mid \omega \in \Omega\}.$$

Similarly, full possibility space Ω is the set that includes all the elements, and therefore it can be seen as an IF-set given by:

$$\Omega = \{(\omega, 1, 0) \mid \omega \in \Omega\}.$$

Definition 2.51 ([6]) *An interval valued fuzzy set is defined by:*

$$A = \{[l_A(\omega), u_A(\omega)] : \omega \in \Omega\},$$

where $0 \leq l_A \leq u_A(\omega) \leq 1$. When $l_A(\omega) = u_A(\omega)$ for any $\omega \in \Omega$, A becomes a fuzzy set with membership function l_A .

If $[l_A(\omega), u_A(\omega)]$ represents that the exact membership degree of ω to A belongs to this interval, the interval $[1 - u_A(\omega), 1 - l_A(\omega)]$ tells us that the exact membership degree of ω to A^c belongs to such interval. Then, A^c is defined by:

$$A^c = \{[1 - u_A(\omega), 1 - l_A(\omega)] : \omega \in \Omega\}.$$

Moreover, the empty set is defined by the interval $[0, 0]$ for any $\omega \in \Omega$, and the total set is defined by the interval $[1, 1]$ for any $\omega \in \Omega$.

IF-sets and IVF-sets are formally equivalent. On the one hand, given an IF-set A with membership and non-membership functions μ_A and ν_A , it defines an IVF-set by:

$$\{[\mu_A(\omega), 1 - \nu_A(\omega)] : \omega \in \Omega\}.$$

On the other hand, given an IVF-set with lower and upper bounds l_A and u_A , it defines an IF-set by:

$$\{(\omega, l_A(\omega), 1 - u_A(\omega)) : \omega \in \Omega\}.$$

For this reason, although the remainder of this section is written in terms of IF-sets, it could be analogously be formulated in terms of IVF-sets.

Let us see how to extend the usual definitions between fuzzy sets, like intersections, unions or differences, towards IF-sets. Similarly to the fuzzy case, unions and intersections of IF-sets are defined by means of t-conorms and t-norms. Recall that a t-norm is a commutative, monotonic and associative binary operator from $[0, 1] \times [0, 1]$ to $[0, 1]$ with neutral element 1, while a t-conorm satisfies the same properties than a t-norm but its neutral element is 0. From a t-norm T it is possible to define a t-conorm S_T , called the dual t-conorm, by:

$$S_T(x, y) = 1 - T(1 - x, 1 - y) \text{ for any } (x, y) \in [0, 1]^2.$$

See [99] for a complete study on t-norms.

Definition 2.52 ([63]) Let A and B be two IF-sets given by:

$$\begin{aligned} A &= \{(\omega, \mu_A(\omega), \nu_A(\omega)) \mid \omega \in \Omega\}. \\ B &= \{(\omega, \mu_B(\omega), \nu_B(\omega)) \mid \omega \in \Omega\}. \end{aligned}$$

Let T be a t-norm and S_T its dual t-conorm.

- The T -intersection of A and B is the IF-set $A \cap_T B$ defined by:

$$A \cap_T B = \{(\omega, T(\mu_A(\omega), \mu_B(\omega)), S_T(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\}.$$

- The S_T -union of A and B is the IF-set $A \cup_{S_T} B$ given by:

$$A \cup_{S_T} B = \{(\omega, S_T(\mu_A(\omega), \mu_B(\omega)), T(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\}.$$

Recall that we shall use the minimum, T_N , and the maximum, S_{T_M} , in order to make intersections and unions, respectively, since they are the most usual operators used in the literature. In that case, the T -intersection and the S_T -union become:

$$\begin{aligned} A \cap_{T_M} B &= \{(\omega, T_M(\mu_A(\omega), \mu_B(\omega)), S_{T_M}(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\} \\ &= \{(\omega, \min(\mu_A(\omega), \mu_B(\omega)), \max(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\}. \\ A \cup_{S_{T_M}} B &= \{(\omega, S_{T_M}(\mu_A(\omega), \mu_B(\omega)), T_M(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\} \\ &= \{(\omega, \max(\mu_A(\omega), \mu_B(\omega)), \min(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\}. \end{aligned}$$

For simplicity, we shall denote the T -intersection and the S_T by \cap and \cup .

We next define a binary relationship of inclusion between IF-sets.

Definition 2.53 *Let A and B be two IF-sets. A is contained in B , and it is denoted by $A \subseteq B$, if*

$$\mu_A(\omega) \leq \mu_B(\omega) \text{ and } \nu_A(\omega) \geq \nu_B(\omega) \text{ for any } \omega \in \Omega.$$

Example 2.54 *Let us consider a possibility space Ω representing a set of three cities: city 1, city 2 and city 3. Let P be a politician, and let us consider the IF-sets:*

$$\begin{aligned} A &= \text{“}P \text{ is a good politician”}. \\ B &= \text{“}P \text{ is honest”}. \\ C &= \text{“}P \text{ is close to the people”}. \end{aligned}$$

Since A , B and C are IF-sets, each city has a degree of agreement with feature A , B and C , and a degree of disagreement. In Figure 2.5 we can see the membership and non-membership functions of these IF-sets.

Now, in order to compute the intersection of the IF-sets A and B ,

$$A \cap B = \text{“}P \text{ is a good politician and honest”}.$$

we must compute the value of $\mu_{A \cap B}$ and

$$\begin{aligned} \mu_{A \cap B}(\text{city } i) &= \min(\mu_A(\text{city } i), \mu_B(\text{city } i)) = \mu_B(\text{city } i), \text{ for } i = 1, 2, 3. \\ \nu_{A \cap B}(\text{city } i) &= \max(\nu_A(\text{city } i), \nu_B(\text{city } i)) = \nu_B(\text{city } i), \text{ for } i = 1, 2, 3. \end{aligned}$$

Thus, $A \cap B = B$. It holds since $B \subseteq A$, in the sense that $\mu_B \leq \mu_A$ and $\nu_B \geq \nu_A$, and its interpretation would be that P is less honest than a good politician.

Now, let us compute the IF-set “ P is honest or close to the people”, that is, the IF-set $B \cup C$. We obtain that:

$$\begin{aligned} \mu_{B \cup C}(\text{city } i) &= \max(\mu_B(\text{city } i), \mu_C(\text{city } i)) = \begin{cases} \mu_B(\text{city } i) & \text{for } i = 1, 3. \\ \mu_C(\text{city } i) & \text{for } i = 2. \end{cases} \\ \nu_{B \cup C}(\text{city } i) &= \min(\nu_B(\text{city } i), \nu_C(\text{city } i)) = \begin{cases} \nu_B(\text{city } i) & \text{for } i = 1, 3. \\ \nu_C(\text{city } i) & \text{for } i = 2. \end{cases} \end{aligned}$$

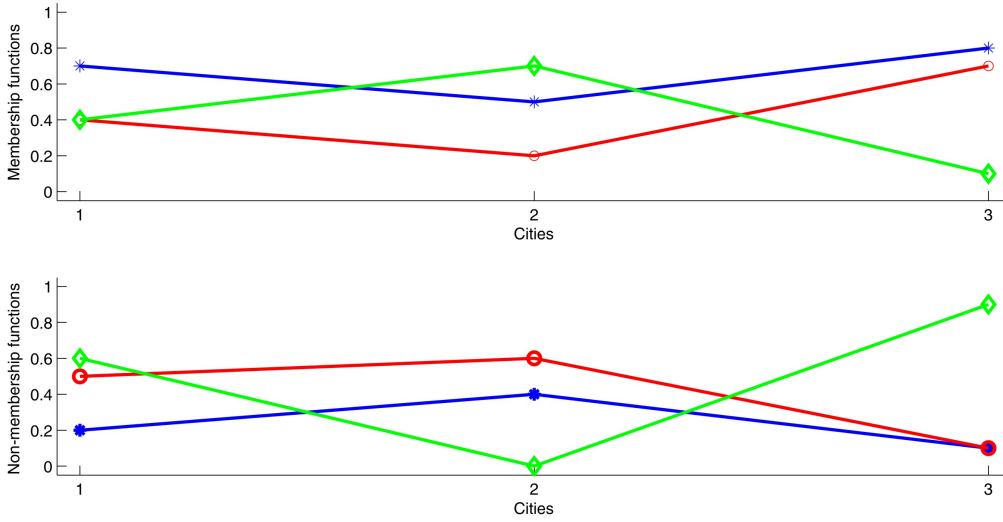


Figure 2.5: Examples of the membership and non-membership functions of the IF-sets that express the P is a good politician (*), P is honest (o) and P is close to the people (◇).

Then, the IF-set $B \cup C$ can be expressed in the following way:

$$B \cup C = \{(city\ 1, \mu_B(city\ 1), \nu_B(city\ 1)), \\ (city\ 2, \mu_C(city\ 2), \nu_C(city\ 2)), (city\ 3, \mu_B(city\ 3), \nu_B(city\ 3))\} \blacklozenge$$

Let us conclude this part by defining the difference operator between IF-sets. According to [27], a difference between fuzzy sets, or *fuzzy difference*, is a map $- : FS(\Omega) \times FS(\Omega) \rightarrow FS(\Omega)$ such that for every pair of fuzzy sets A and B it satisfies the following properties:

$$\begin{aligned} \text{If } A \subseteq B, \text{ then } A - B &= \emptyset. \\ \text{If } A \subseteq A', \text{ then } A - B &\subseteq A' - B. \end{aligned}$$

Some examples of fuzzy differences are the following:

$$\begin{aligned} A - B(\omega) &= \max\{0, A(\omega) - B(\omega)\}, \\ A - B(\omega) &= \begin{cases} A(\omega) & \text{if } B(\omega) = 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

for any $\omega \in \Omega$.

Similarly, we can extend the definition of difference for IF-sets.

Definition 2.55 An operator $- : IFSS(\Omega) \times IFSS(\Omega) \rightarrow IFSS(\Omega)$ is a difference between IF-sets (IF-difference, in short) if it satisfies properties D1 and D2.

D1 If $A \subseteq B$, then $A - B = \emptyset$.

D2 If $A \subseteq A'$, then $A - B \subseteq A' - B$.

Any function D satisfying D1 and D2 is a difference operator. Nevertheless, there are other interesting properties that IF-differences may satisfy:

D3 $(A \cap C) - (B \cap C) \subseteq A - B$.

D4 $(A \cup C) - (B \cup C) \subseteq A - B$.

D5 $A - B = \emptyset \Rightarrow A \subseteq B$.

Let us give an example of IF-difference that also fulfills D3, D4 and D5.

Example 2.56 Consider the function $- : IFSS(\Omega) \times IFSS(\Omega) \rightarrow IFSS(\Omega)$ given by:

$$A - B = \{(\omega, \mu_{A-B}(\omega), \nu_{A-B}(\omega)) \mid \omega \in \Omega\},$$

where

$$\begin{aligned} \mu_{A-B}(\omega) &= \max(0, \mu_A(\omega) - \mu_B(\omega)); \\ \nu_{A-B}(\omega) &= \begin{cases} 1 - \mu_{A-B}(\omega) & \text{if } \nu_A(\omega) > \nu_B(\omega); \\ \min(1 + \nu_A(\omega) - \nu_B(\omega), 1 - \mu_{A-B}(\omega)) & \text{if } \nu_A(\omega) \leq \nu_B(\omega). \end{cases} \end{aligned}$$

Let us prove that this function satisfies properties D1 and D2, i.e., that it is an IF-difference.

D1: Let us take $A \subseteq B$. Then $\mu_A \leq \mu_B$ and $\nu_A \geq \nu_B$.

$$\begin{aligned} \mu_{A-B}(\omega) &= \max(0, \mu_A(\omega) - \mu_B(\omega)) = 0. \\ \nu_{A-B}(\omega) &= 1 - \mu_{A-B}(\omega) = 1, \text{ because } \nu_A \geq \nu_B. \end{aligned}$$

As a consequence, $A - B = \emptyset$.

D2: Consider $A \subseteq A'$, that is, $\mu_A \leq \mu_{A'}$ and $\nu_A \geq \nu_{A'}$, and let us prove that $A - B \subseteq A' - B$. Thus, for any ω in Ω we have that:

$$\begin{aligned} \mu_{A-B}(\omega) &= \max(0, \mu_A(\omega) - \mu_B(\omega)) \leq \max(0, \mu_{A'}(\omega) - \mu_B(\omega)) = \mu_{A'-B}(\omega). \\ \nu_{A'-B}(\omega) &= \begin{cases} 1 - \mu_{A'-B}(\omega) & \text{if } \nu_{A'}(\omega) > \nu_B(\omega). \\ \min(1 - \mu_{A'-B}(\omega), 1 + \nu_{A'}(\omega) - \nu_B(\omega)) & \text{if } \nu_{A'}(\omega) \leq \nu_B(\omega). \end{cases} \\ &\leq \begin{cases} 1 - \mu_{A-B}(\omega) & \text{if } \nu_{A'}(\omega) > \nu_B(\omega). \\ \min(1 - \mu_{A-B}(\omega), 1 + \nu_A(\omega) - \nu_B(\omega)) & \text{if } \nu_{A'}(\omega) \leq \nu_B(\omega). \end{cases} \\ &\leq \nu_{A-B}(\omega). \end{aligned}$$

This shows that $-$ is an IF-difference. Let us see that it also satisfies properties D3, D4 and D5.

D3: Let us take into account that the IF-sets $A \cap C$ and $B \cap C$ are given by:

$$\begin{aligned} A \cap C &= \{(\omega, \min(\mu_A(\omega), \mu_C(\omega)), \max(\nu_A(\omega), \nu_C(\omega))) \mid \omega \in \Omega\} \\ B \cap C &= \{(\omega, \min(\mu_B(\omega), \mu_C(\omega)), \max(\nu_B(\omega), \nu_C(\omega))) \mid \omega \in \Omega\}. \end{aligned}$$

For short, we will denote by D the IF-set $D = A \cap C - B \cap C$. On one hand, we are going to prove that $\mu_{A-B} \geq \mu_D$:

$$\begin{aligned} \mu_{A-B}(\omega) &= \max(0, \mu_A(\omega) - \mu_B(\omega)). \\ \mu_D(\omega) &= \max(0, \min(\mu_A(\omega), \mu_C(\omega)) - \min(\mu_B(\omega), \mu_C(\omega))). \end{aligned}$$

Applying the first part of Lemma A.1 of Appendix A, we deduce that $\mu_{A-B} \geq \mu_D$.

Now, let us prove that $\nu_{A-B} \leq \nu_D$. There are two possibilities, either $\nu_A(\omega) > \nu_B(\omega)$ or $\nu_A(\omega) \leq \nu_B(\omega)$. Assume that $\nu_A(\omega) > \nu_B(\omega)$. In such a case, $\max(\nu_A(\omega), \nu_C(\omega)) \geq \max(\nu_B(\omega), \nu_C(\omega))$, and $\nu_{A-B}(\omega) = 1 - \mu_{A-B}(\omega)$, and consequently:

$$\nu_D(\omega) = 1 - \mu_D(\omega) \geq 1 - \mu_{A-B}(\omega) = \nu_{A-B}(\omega).$$

Assume now that $\nu_A(\omega) \leq \nu_B(\omega)$. Then it holds that

$$\max(\nu_A(\omega), \nu_C(\omega)) \leq \max(\nu_B(\omega), \nu_C(\omega)).$$

By the second part of Lemma A.1 of Appendix A,

$$\nu_B(\omega) - \nu_A(\omega) \geq \max(\nu_B(\omega), \nu_C(\omega)) - \max(\nu_A(\omega), \nu_C(\omega)),$$

whence

$$\begin{aligned} \nu_D(\omega) &= \min(1 + \max(\nu_A(\omega), \nu_C(\omega)) - \max(\nu_B(\omega), \nu_C(\omega)), 1 - \mu_D(\omega)) \\ &\geq \min(1 + \nu_A(\omega) - \nu_B(\omega), 1 - \mu_{A-B}(\omega)) = \nu_{A-B}(\omega). \end{aligned}$$

Thus we conclude that $\nu_{A-B} \leq \nu_D$, and therefore $(A \cap C) - (B \cap C) \subseteq A - B$.

D4: Consider three IF-sets A, B and C . The IF-sets $A \cup C$ and $B \cup C$ are given by:

$$\begin{aligned} A \cup C &= \langle \max(\mu_A, \mu_C), \min(\nu_A, \nu_C) \rangle. \\ B \cup C &= \langle \max(\mu_B, \mu_C), \min(\nu_B, \nu_C) \rangle. \end{aligned}$$

Let us denote by D the IF-set $D = (A \cup C) - (B \cup C)$, and let us prove that $\mu_{A-B} \geq \mu_D$. This is equivalent to

$$\max(0, \mu_A(\omega) - \mu_B(\omega)) \geq \max(0, \max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))),$$

for every $\omega \in \Omega$, and this inequality holds because of the first part of Lemma A.1 of Appendix A.

Let us prove that $\nu_D \geq \nu_{A-B}$. To see this, consider the two possible cases: $\nu_A(\omega) > \nu_B(\omega)$ and $\nu_A(\omega) \leq \nu_B(\omega)$. Assume that $\nu_A(\omega) > \nu_B(\omega)$, which means that $\nu_{A-B}(\omega) =$

$1 - \mu_{A-B}(\omega)$. Now, $\nu_A(\omega) > \nu_B(\omega)$ implies that $\min(\nu_A(\omega), \nu_C(\omega)) \geq \min(\nu_B(\omega), \nu_C(\omega))$, and therefore:

$$\nu_D(\omega) = 1 - \mu_D(\omega) \geq 1 - \mu_{A-B}(\omega) = \nu_{A-B}(\omega).$$

Assume now that $\nu_A(\omega) \leq \nu_B(\omega)$, whence

$$\min(\nu_A(\omega), \nu_C(\omega)) \leq \min(\nu_B(\omega), \nu_C(\omega)).$$

Applying the second part of Lemma A.1 of Appendix A, we know that

$$\nu_B(\omega) - \nu_A(\omega) \geq \min(\nu_B(\omega), \nu_C(\omega)) - \min(\nu_A(\omega), \nu_C(\omega)).$$

Then, we deduce that:

$$\begin{aligned} \nu_D(\omega) &= \min(1 + \min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega)), 1 - \mu_D(\omega)) \\ &\geq \min(1 + \nu_A(\omega) - \nu_B(\omega), 1 - \mu_A(\omega)) = \nu_{A-B}(\omega). \end{aligned}$$

Thus, $\nu_D \geq \nu_{A-B}$, and therefore $(A \cup C) - (B \cup C) \subseteq A - B$.

D5: Let us consider A and B such that $A - B = \emptyset$. Then, $\mu_{A-B}(\omega) = 0$ and $\nu_{A-B}(\omega) = 1$ for every $\omega \in \Omega$, whence

$$\begin{aligned} 0 = \mu_{A-B}(\omega) &= \max(0, \mu_A(\omega) - \mu_B(\omega)) \Rightarrow \mu_A(\omega) \leq \mu_B(\omega). \\ 1 = \nu_{A-B}(\omega) &= \begin{cases} 1 & \text{if } \nu_A(\omega) > \nu_B(\omega). \\ 1 + \nu_A(\omega) - \nu_B(\omega) & \text{if } \nu_A(\omega) \leq \nu_B(\omega). \end{cases} \end{aligned}$$

Therefore, $\mu_A(\omega) \leq \mu_B(\omega)$ and $\nu_A(\omega) \geq \nu_B(\omega)$, and as a consequence $A \subseteq B$. ♦

3 Comparison of alternatives under uncertainty

This memory is devoted to the comparison of alternatives under some lack of information. If this lack of information is given by uncertainty about the consequences of the alternatives, these are usually modelled by means of random variables. Thus, stochastic orders emerge as an essential tool, since they allow the comparison of random quantities. As we mentioned in the previous chapter, one of the most important stochastic orders in the literature is that of stochastic dominance, in any of its degrees. Stochastic dominance has been widely investigated (see [98, 108, 109, 110, 173], among others) and it has been applied in many different areas ([11, 77, 95, 109, 171, 180]). However, the other stochastic order we have introduced, statistical preference, has been studied in ([14, 15, 16, 49, 54, 55, 56, 57, 58]) but not as widely as stochastic dominance. For this reason, the first step of this chapter is to make a thorough study of statistical preference. First of all, we investigate its basic properties as a stochastic order, and then we study its relationship with stochastic dominance. In this sense, we shall firstly look for conditions that guarantee that first degree stochastic dominance implies statistical preference. Then, we shall show that in general there is not an implication relationship between statistical preference and the n -th degree stochastic dominance. We also provide several examples of the behaviour of statistical preference, and also stochastic dominance, in some of the most usual distributions, like for instance Bernoulli, exponential or, of course, the normal distribution.

Both stochastic dominance and statistical preference are stochastic orders that were introduced for the pairwise comparison of random variables. In fact, statistical preference presents a disadvantage that is its lack of transitivity, as was pointed out by several authors ([14, 15, 16, 49, 54, 56, 58, 121, 122]). To illustrate this fact, we give an example. Then, in order to have an stochastic order that allows for the simultaneous comparison of more than two random variables, we present a generalisation of statistical preference, and study some of its properties. In particular, we shall see its connections with the methods established for pairwise comparisons.

It is obvious that stochastic orders are powerful tools for comparing uncertain quan-

tities. For this reason, and in order to illustrate our results, we conclude the chapter by mentioning two possible applications. On the one hand, we investigate both stochastic dominance and statistical preference as methods for the comparison of fitness values ([180, 183]), and on the other hand we illustrate the usefulness of both statistical preference and its generalisation for the comparison of more than two random variables in multicriteria decision making problems with linguistic labels ([123]).

3.1 Properties of the statistical preference

This section is devoted to the study of the main properties of statistical preference. In particular, we shall try to find a characterization of this notion: on a first step, a similar one to that of stochastic dominance presented in Theorem 2.10; afterwards, we explain that statistical preference seems to be closer to another location parameter, the median.

3.1.1 Basic properties and intuitive interpretation of the statistical preference

We start this subsection with some basic properties about the behaviour of the statistical preference.

Lemma 3.1 *Let X and Y be two random variables. Then it holds that*

$$X \succeq_{\text{SP}} Y \Rightarrow P(X < Y) \leq \frac{1}{2}.$$

In particular, the converse implication holds for random variables with $P(X = Y) = 0$.

Proof: It holds that $Q(X, Y) = P(X > Y) + \frac{1}{2}P(X = Y) \geq \frac{1}{2}$. Then:

$$P(X < Y) = 1 - P(X > Y) - P(X = Y) \leq \frac{1}{2} - \frac{1}{2}P(X = Y) \leq \frac{1}{2}.$$

If $P(X = Y) = 0$, then:

$$Q(X, Y) = P(X > Y) = 1 - P(Y > X) \geq \frac{1}{2},$$

since we assume $P(X < Y) \leq \frac{1}{2}$. Thus, $X \succeq_{\text{SP}} Y$. ■

Remark 3.2 *Note that the converse implication of the previous result does not hold in general. As a counterexample, it is enough to consider the independent random variables*

defined by:

$$\frac{X}{P_X} \begin{array}{c|cc} & 0 & 2 \\ \hline & 0.8 & 0.2 \end{array} \quad \frac{Y}{P_Y} \begin{array}{c|cc} & 0 & 1 \\ \hline & 0.7 & 0.3 \end{array}$$

On the one hand, it holds that:

$$P(X < Y) = P(X = 0, Y = 1) = P(X = 0)P(Y = 1) = 0.8 \cdot 0.3 = 0.24 < \frac{1}{2},$$

and also

$$P(X = Y) = P(X = 0, Y = 0) = P(X = 0)P(Y = 0) = 0.7 \cdot 0.8 = 0.56.$$

However, $P(X > Y) = P(X = 2) = 0.2$. Thus:

$$Q(X, Y) = P(X > Y) + \frac{1}{2}P(X = Y) = 0.2 + \frac{1}{2} \cdot 0.56 = 0.48 < \frac{1}{2}. \blacklozenge$$

Now we present a result that shows how translations and dilations or contractions affect to the behaviour of statistical preference for real-valued random variables.

Proposition 3.3 *Let X , Y and Z be three real-valued random variables defined on the same probability space and let $\lambda \neq 0$ and μ be two real numbers. It holds that*

1. $X \succeq_{\text{SP}} Y \Leftrightarrow X + Z \succeq_{\text{SP}} Y + Z$.
2. $\lambda X \succeq_{\text{SP}} \mu Y \Leftrightarrow \begin{cases} X \succeq_{\text{SP}} \frac{\mu}{\lambda} Y & \text{if } \lambda > 0. \\ \frac{\mu}{\lambda} Y \succeq_{\text{SP}} X & \text{if } \lambda < 0. \end{cases}$

Proof:

1. It holds that

$$\begin{aligned} Q(X, Y) &= P(X > Y) + \frac{1}{2}P(X = Y) \\ &= P(X + Z > Y + Z) + \frac{1}{2}P(X + Z = Y + Z) = Q(X + Z, Y + Z). \end{aligned}$$

Then, $Q(X, Y) \geq \frac{1}{2}$ if and only if $Q(X + Z, Y + Z) \geq \frac{1}{2}$.

2. Let us develop the expression of $Q(\lambda X, \mu Y)$:

$$Q(\lambda X, \mu Y) = \begin{cases} P(X > \frac{\mu}{\lambda} Y) + P(X = \frac{\mu}{\lambda} Y) = Q(X, \frac{\mu}{\lambda} Y) & \text{if } \lambda > 0. \\ P(X < \frac{\mu}{\lambda} Y) + P(X = \frac{\mu}{\lambda} Y) = Q(\frac{\mu}{\lambda} Y, X) & \text{if } \lambda < 0. \end{cases}$$

Then, the result direct follows from the expression of $Q(\lambda X, \mu Y)$. ■

Some new equivalences can be deduced from the previous ones.

Corollary 3.4 *Let X and Y be a pair of real-valued random variables, λ and μ two real numbers and α a constant. Then it holds that*

1. $\lambda X \succeq_{\text{SP}} \mu \Leftrightarrow \begin{cases} X \succeq_{\text{SP}} \frac{\mu}{\lambda}, & \text{if } \lambda > 0, \\ \frac{\mu}{\lambda} \succeq_{\text{SP}} X, & \text{if } \lambda < 0, \\ 0 \geq \mu, & \text{if } \lambda = 0. \end{cases}$
2. $X \succeq_{\text{SP}} Y \Leftrightarrow 1 - Y \succeq_{\text{SP}} 1 - X.$
3. $X \succeq_{\text{SP}} Y \Leftrightarrow X - Y \succeq_{\text{SP}} 0.$
4. $X + Y \succeq_{\text{SP}} Y \Leftrightarrow X \succeq_{\text{SP}} 0.$
5. $X \succeq_{\text{SP}} X + \alpha \Leftrightarrow \alpha \leq 0.$
6. $X \succeq_{\text{SP}} \alpha X \Leftrightarrow \begin{cases} 0 \succeq_{\text{SP}} X, & \text{if } \alpha > 1, \\ X \succeq_{\text{SP}} 0, & \text{if } \alpha < 1. \end{cases}$

Proof: In point 1, the case of $\lambda > 0$ and $\lambda < 0$ directly follow from item 2 of the previous proposition. If $\lambda = 0$, applying Remark 2.19, the comparison of degenerate random variables is equivalent to the comparison of real numbers, and then, it is obvious that $\lambda X \succeq_{\text{SP}} \mu \Leftrightarrow 0 \geq \mu$.

Point 2 follows from the previous proposition: $X \succeq_{\text{SP}} Y$ if and only if $X - 1 \succeq_{\text{SP}} Y - 1$, and from the second item this is equivalent to $1 - Y \succeq_{\text{SP}} 1 - X$.

Points 3, 4 and 5 are immediate from the first point of Proposition 3.3 and Remark 2.19 in the case of 3. Consider the last one. Applying our previous proposition,

$$X \succeq_{\text{SP}} \alpha X \Leftrightarrow (1 - \alpha)X \succeq_{\text{SP}} 0.$$

By the second item of Proposition 3.3,

$$(1 - \alpha)X \succeq_{\text{SP}} 0 \Leftrightarrow \begin{cases} 0 \succeq_{\text{SP}} X, & \text{if } \alpha > 1, \\ X \succeq_{\text{SP}} 0, & \text{if } \alpha < 1. \end{cases} \quad \blacksquare$$

Let us compare the behaviour of statistical preference and stochastic dominance with respect these basic properties. On the one hand, Proposition 2.13 assures that $X_1 + \dots + X_n \succeq_{\text{FSD}} Y_1 + \dots + Y_n$ when the variables are independent and $X_i \succeq_{\text{FSD}} Y_i$. First statement of Proposition 3.3 assures that $X \succeq_{\text{SP}} Y \Leftrightarrow X + Z \succeq_{\text{SP}} Y + Z$, and the independence condition is not imposed. However, it is not possible to give a result as general as Proposition 2.13 for statistical preference. For instance, consider the universe

$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, with a discrete uniform distribution, and the following random variables:

	ω_1	ω_2	ω_3	ω_4
X_1	-2	1	-2	1
X_2	1	-2	-2	1
Y	0	0	0	0
$X_1 + X_2$	-1	-1	-4	2
$Y + Y$	0	0	0	0

It holds that $X_1 \equiv_{\text{SP}} Y$ and $X_2 \equiv_{\text{SP}} Y$. However, $Q(X_1 + X_2, Y + Y) = \frac{1}{4}$, and therefore $X_1 + X_2 \not\equiv_{\text{SP}} Y + Y$.

First item of Corollary 3.4 trivially holds for stochastic dominance. The second item also holds since:

$$F_{1-X}(t) = 1 - P(X < 1 - t) \text{ and } F_{1-Y}(t) = 1 - P(Y < 1 - t),$$

and then $F_{1-Y}(t) \leq F_{1-X}(t)$ if and only if $P(X < 1 - t) \leq P(Y < 1 - t)$. Note that $P(X \leq t) \leq P(Y \leq t)$ for any t if and only if $P(X < t) \leq P(Y < t)$ for any t : on the one hand, assume that $P(X \leq t) \leq P(Y \leq t)$ for any t . Then:

$$P(X < t') = \lim_{n \rightarrow \infty} P\left(X \leq t' - \frac{1}{n}\right) \leq \lim_{n \rightarrow \infty} P\left(Y \leq t' - \frac{1}{n}\right) = P(Y < t').$$

On the other hand, if $P(X < t) \leq P(Y < t)$ for any t , it holds that:

$$P(X \leq t') = \lim_{n \rightarrow \infty} P\left(X < t' + \frac{1}{n}\right) \leq \lim_{n \rightarrow \infty} P\left(Y < t' + \frac{1}{n}\right) = P(Y \leq t').$$

We conclude that $X \succeq_{\text{FSD}} Y$ if and only if $1 - Y \succeq_{\text{FSD}} 1 - X$. However, stochastic dominance does not satisfy the third item of Corollary 3.4. For instance, if X and Y are two independent and equally distributed random variables following a Bernoulli distribution of parameter $\frac{1}{2}$, it holds that:

$X - Y$	-1	0	1
P_{X-Y}	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Then, $X - Y$ is not comparable with the degenerate variable in 0 for first degree stochastic dominance, but $X \succeq_{\text{FSD}} Y$.

Furthermore, the fourth item of the previous corollary does not hold, either: it suffices to consider the universe $\Omega = \{\omega_1, \omega_2, \omega_3\}$ with discrete uniform distribution, and the random variables defined by:

	ω_1	ω_2	ω_3
X	0	1	2
Y	2	1	0
$X - Y$	-2	0	2

Then, $X \equiv_{\text{FSD}} Y$, but $X - Y$ and 0 are not comparable with respect to stochastic dominance. Nevertheless, first degree stochastic dominance does satisfy the fifth and sixth properties of Corollary 3.4.

Remark 3.5 *Using the third item of the previous corollary, we know that $X \succeq_{\text{SP}} Y$ if and only if $X - Y \succeq_{\text{SP}} 0$. This allowed Couso and Sánchez [46] to prove a simple characterization of statistical preference for real-valued random variables:*

$$X \succeq_{\text{SP}} Y \Leftrightarrow X - Y \succeq_{\text{SP}} 0 \Leftrightarrow E[u(X - Y)] \geq 0 \quad (3.1)$$

for the function $u : \mathbb{R} \rightarrow \mathbb{R}$ defined by $u = I_{(0,\infty)} - I_{(-\infty,0)}$.♦

Theorem 2.10 showed that $X \succeq_{\text{FSD}} Y$ if and only if the expectation of $u(X)$ is greater than the expectation of $u(Y)$ for any increasing function u . In particular, Proposition 2.12 assures that, when $X \succeq_{\text{FSD}} Y$ and φ is a increasing function, $\varphi(X) \succeq_{\text{FSD}} \varphi(Y)$. In the case of statistical preference, we can check that it is invariant by strictly increasing transformations of the random variables as well.

Proposition 3.6 *Let X and Y be two random variables. It holds that:*

$$X \succeq_{\text{SP}} Y \Leftrightarrow h(X) \succeq_{\text{SP}} h(Y)$$

for any strictly order preserving function $h : \Omega' \rightarrow \Omega'$.

Proof: On the one hand, if $h(X) \succeq_{\text{SP}} h(Y)$ for any strictly order preserving function h , by considering the identity function we obtain that $X \succeq_{\text{SP}} Y$.

On the other hand, note that:

$$\{\omega : h(X(\omega)) > h(Y(\omega))\} = \{\omega : X(\omega) > Y(\omega)\},$$

and consequently $P(X > Y) = P(h(X) > h(Y))$. Similarly, $P(X = Y) = P(h(X) = h(Y))$ and $P(Y > X) = P(h(Y) > h(X))$. Then $Q(X, Y) = Q(h(X), h(Y))$. We conclude that $X \succeq_{\text{SP}} Y \Leftrightarrow h(X) \succeq_{\text{SP}} h(Y)$. ■

However, although first degree stochastic dominance is invariant under increasing transformations, for statistical preference the previous result does not hold for order preserving functions that are not strictly order preserving. For instance, consider the following independent random variables:

$$\frac{X}{P_X} \begin{array}{c|cc} & 0 & 2 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \quad \frac{Y}{P_Y} \begin{array}{c|c} & 1 \\ \hline & 1 \end{array}$$

Then, the probabilistic relation takes the value $Q(X, Y) = \frac{1}{2}$. Consider the increasing, but not strictly increasing, function $h : \mathbb{R} \rightarrow \mathbb{R}$ given by:

$$h(t) = \begin{cases} t & \text{if } t \in (-\infty, 0] \cup (2, \infty). \\ 2 & \text{otherwise.} \end{cases}$$

Then, $h(X)$ and $h(Y)$ are given by:

$$\frac{h(X)}{P_{h(X)}} \left| \begin{array}{c|c} 0 & 2 \\ \hline \frac{1}{2} & \frac{1}{2} \end{array} \right. \quad \frac{h(Y)}{P_{h(Y)}} \left| \begin{array}{c|c} 2 & \\ \hline 1 & \end{array} \right.$$

Thus, $Q(h(X), h(Y)) = \frac{1}{4}$, and then the previous result does not hold.

The last basic property we are going to study is if statistical preference is preserved by different kinds of convergence.

Remark 3.7 *Let $\{X_n\}_n$ and $\{Y_n\}_n$ be two sequences of random variables and let X and Y other two random variables, all of them defined on the same probability space. It holds that:*

$$\left. \begin{array}{l} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} Y \\ X_n \succeq_{\text{SP}} Y_n \quad \forall n \end{array} \right\} \not\Rightarrow X \succeq_{\text{SP}} Y,$$

$$\left. \begin{array}{l} X_n \xrightarrow{\mathcal{P}} X \\ Y_n \xrightarrow{\mathcal{P}} Y \\ X_n \succeq_{\text{SP}} Y_n \quad \forall n \end{array} \right\} \not\Rightarrow X \succeq_{\text{SP}} Y,$$

$$\left. \begin{array}{l} X_n \xrightarrow{m-p} X \\ Y_n \xrightarrow{m-p} Y \\ X_n \succeq_{\text{SP}} Y_n \quad \forall n \end{array} \right\} \not\Rightarrow X \succeq_{\text{SP}} Y,$$

$$\left. \begin{array}{l} X_n \xrightarrow{\text{a.s.}} X \\ Y_n \xrightarrow{\text{a.s.}} Y \\ X_n \succeq_{\text{SP}} Y_n \quad \forall n \end{array} \right\} \not\Rightarrow X \succeq_{\text{SP}} Y,$$

where $\xrightarrow{\mathcal{L}}$, $\xrightarrow{\mathcal{P}}$, $\xrightarrow{m-p}$ and $\xrightarrow{\text{a.s.}}$ denote the convergence of random variables in distribution, probability, m^{th} -mean and almost sure, respectively.

It suffices to consider the same counterexample for all the cases: consider the universe $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and the probability P such that $P(\{\omega_1\}) = P(\{\omega_3\}) = \frac{2}{5}$ and $P(\{\omega_2\}) = P(\{\omega_4\}) = \frac{1}{10}$. Let X , X_n , Y and Y_n be the random variables defined by:

	ω_1	ω_2	ω_3	ω_4
X, X_n	0	0	1	1
Y	0	1	1	1
Y_n	$\frac{-1}{n}$	1	1	1

Y_n converges to Y almost surely, and consequently also converges in probability and in distribution. Furthermore, it also converges in m^{th} mean, since:

$$E[|Y_n - Y|^m] = \frac{2}{5} \left(\frac{1}{n}\right)^m \xrightarrow{n \rightarrow \infty} 0.$$

Also, X_n converges to X for the four kinds of convergence. Furthermore, $X_n \succ_{\text{SP}} Y_n$ since:

$$\begin{aligned} Q(X_n, Y_n) &= P(X_n > Y_n) + \frac{1}{2}P(X_n = Y_n) \\ &= P(\{\omega_1\}) + \frac{1}{2}P(\{\omega_3, \omega_4\}) = \frac{2}{5} + \frac{1}{2} \cdot \frac{1}{2} = \frac{13}{20} > \frac{1}{2}. \end{aligned}$$

However, $X \not\succeq_{\text{SP}} Y$, since:

$$Q(X, Y) = P(X > Y) + \frac{1}{2}P(X = Y) = \frac{1}{2}P(\{\omega_1, \omega_3, \omega_4\}) = \frac{9}{20} < \frac{1}{2}. \blacklozenge$$

Thus, we can see that, although stochastic dominance is preserved by the four kind of convergence (see Prop. 2.14), statistical preference is not.

Now we shall try to clarify the meaning of statistical preference by means of a gambling example.

Example 3.8 Suppose we have two random variables X and Y defined over the same probability space such that $X \succ_{\text{SP}} Y$, i.e., such that $Q(X, Y) > \frac{1}{2}$. Consider the following game: we obtain a pair of random values of X and Y simultaneously. For example, if X and Y model the results of the dice, we would roll them simultaneously; otherwise, they can be simulated by a computer. Player 1 bets 1 euro on Y to take a value greater than X . If this holds, Player 1 wins 1 euro, he loses 1 euro if the value of X is greater, and he does not lose anything if the values are equal.

Denote by Z_i the random variable “reward of Player 1 in the i -th iteration of the game”. Then it holds that

$$Z_i = \begin{cases} 1, & \text{if } Y > X \\ 0, & \text{if } Y = X \\ -1, & \text{if } Y < X \end{cases}$$

Then, applying the hypothesis $P(X > Y) + \frac{1}{2}P(X = Y) > \frac{1}{2}$, it holds that

$$\begin{aligned} P(X > Y) &> \frac{1}{2}(1 - P(X = Y)) = \frac{1}{2}(P(X > Y) + P(Y > X)) \\ &\Rightarrow P(X > Y) > P(Y > X), \end{aligned}$$

or equivalently, $q > p$, if we consider the notation $p = P(X < Y)$ and $q = P(X > Y)$. Thus

$$E(Z_i) = P(Y > X) - P(Y < X) = p - q < 0.$$

$\{Z_1, Z_2, \dots\}$ is an infinite sequence of independent and identically distributed random variables. Applying the large law of big numbers,

$$\overline{Z_n} = \frac{Z_1 + \dots + Z_n}{n} \xrightarrow{p} p - q,$$

or equivalently,

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|\bar{Z}_n - (p - q)| > \varepsilon) = 0. \quad (3.2)$$

Denote the accumulated reward of Player 1 after n iterations of the game by S_n . It holds that $S_n = Z_1 + \dots + Z_n$. Then, Player 1 wins the game after n iterations if $S_n > 0$. Then, taking $\varepsilon = q - p$ in Equation (3.2), Player 1 wins the game after n iterations with probability:

$$\begin{aligned} P(S_n > 0) &= P(Z_1 + \dots + Z_n > 0) = P(\bar{Z}_n > 0) \\ &= P(\bar{Z}_n - (p - q) > q - p) \leq P(|\bar{Z}_n - (p - q)| > q - p) \\ &= P(|\bar{Z}_n - (p - q)| > \varepsilon). \end{aligned}$$

Then it holds that:

$$\lim_{n \rightarrow \infty} P(S_n > 0) \leq \lim_{n \rightarrow \infty} P(|\bar{Z}_n - (p - q)| > \varepsilon) = 0.$$

We have proven that the probability of the event: “Player 1 wins after n iterations of the game” goes to 0 when n goes to ∞ . ♦

An immediate consequence is the next proposition:

Proposition 3.9 *Let X and Y be two random variables such that $X \succ_{\text{SP}} Y$. Consider the experiment that consists of drawing a random sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of X and Y , and let*

$$B_n \equiv \text{“In the first } n \text{ iterations, at least half of the times the value obtained by } X \text{ is greater than or equal to the value obtained by } Y\text{”}.$$

Then,

$$\lim_{n \rightarrow \infty} P(B_n) = 1.$$

Then we can say that if we consider the game consisting of obtaining a random value of X and a random value of Y and we repeat it a large enough number of times, if $X \succ_{\text{SP}} Y$, we will obtain that more than half of the times the variable X will take a value greater than the value obtained by Y . However, this does not guarantee that the mean value obtained by the variable X is greater than the mean value obtained by the variable Y .

Let us consider a new example:

Example 3.10 ([57]) *Let us consider the game consisting of rolling two special dice, denoted A and B , whose results are assumed to be independent. Their faces do not show the classical values but the following numbers:*

<u>DICE A</u>			
	1		
3	4	15	16
	17		

<u>DICE B</u>			
	2		
10	11	12	13
	14		

In each iteration, the dice with the greatest number wins.

In this case the probabilistic relation Q of Equation (2.7) takes the value:

$$Q(A, B) = P(A > B) + \frac{1}{2}P(A = B) = P(A > B) = P(A \in \{3, 4\}, B \in \{2\}) \\ + P(A \in \{15, 16, 17\}, B \in \{2, 10, 11, 12, 13, 14\}) = \frac{5}{9}.$$

Thus, $A \succ_{\text{SP}} B$ and applying the previous result, if we repeat the game indefinitely, it holds that the probability of winning, betting on A , at least half of the times tends to 1.

However, if we calculate the expected value of every dice, we obtain that

$$E(A) = \frac{1}{6}(1 + 3 + 4 + 15 + 16 + 17) = \frac{28}{3}, \\ E(B) = \frac{1}{6}(2 + 10 + 11 + 12 + 13 + 14) = \frac{31}{3}.$$

Then, by the criterium of the highest expected reward, dice B should be preferred. The same applies if we consider the criterion of stochastic dominance. However, if our goal is to win the majority of times then we should choose dice A . ♦

3.1.2 Characterizations of statistical preference

In Subsection 2.1.1 we have seen that stochastic dominance can be characterised by means of the direct comparison of the expectation of adequate transformations of the random variables (see Theorem 2.10). In this subsection we shall give characterisations for statistical preference. For this aim, we distinguish different cases: we start by considering independent random variables, then we consider comonotonic and countermonotonic random variables and we conclude with random variables coupled by means of an Archimedean copula. Finally, we show an alternative characterization of statistical preference in terms of the median. Recall that in the rest of this section, we will consider real-valued random variables.

Independent random variables

We start by considering independent random variables. In order to characterise statistical preference for them, we need this previous result.

Lemma 3.11 *Consider two independent real-valued random variables X and Y whose associated cumulative distribution functions are F_X and F_Y , respectively. Then:*

$$P(X \geq Y) = E[F_Y(X)], \quad (3.3)$$

where $E[h(X)]$ stands for the expectation of the function h with respect to the variable X , this is, $E[h(X)] = \int h(x)dF_X(x)$.

Proof: In order to prove this result, we consider [24, Theorem 20.3]: given two random vectors X and Y defined on \mathbb{R}^j and \mathbb{R}^k , and whose distribution functions are F_X and F_Y , respectively, it holds that:

$$P((X, Y) \in B) = \int_{\mathbb{R}^j} P((x, Y) \in B)dF_X(x), \quad B \in \mathbb{R}^{j+k}. \quad (3.4)$$

In this case, consider $j = k = 1$ and $B = \{(x, y) : x \geq y\}$. Then:

$$\begin{aligned} P((X, Y) \in B) &= P(X \geq Y) \text{ and} \\ P((x, Y) \in B) &= P(Y \leq x) = F_Y(x). \end{aligned}$$

Then, if we put these values into Equation (3.4), we obtain that $P(X \geq Y) = E[F_Y(X)]$. ■

We can now establish the following result.

Theorem 3.12 *Let X and Y be two independent real-valued random variables defined on the same probability space. Let F_X and F_Y denote their respective cumulative distribution functions. If X' is a random variable identically distributed to X and independent of X and Y , it holds that $X \succeq_{\text{SP}} Y$ if and only if:*

$$E[F_Y(X)] - E[F_X(X)] \geq \frac{1}{2}(P(X = Y) - P(X = X')). \quad (3.5)$$

Proof: It holds that $X \succeq_{\text{SP}} Y$ if and only if $P(Y > X) + \frac{1}{2}P(X = Y) \leq \frac{1}{2}$. On the other hand let us recall (see for example [24, Exercise 21.9(d)]) that $E(F_X(X)) = \frac{1}{2} + \frac{1}{2}P(X = X')$. Then, using also Equation (3.3):

$$\begin{aligned} P(Y > X) &= 1 - P(Y \leq X) = 1 - E[F_Y(X)] \\ &= \frac{1}{2} + E[F_X(X)] - \frac{1}{2}P(X = X') - E[F_Y(X)]. \end{aligned}$$

Whereas, $X \succeq_{\text{SP}} Y$ if and only if

$$\frac{1}{2} + E[F_X(X)] - \frac{1}{2}P(X = X') - E[F_Y(X)] + \frac{1}{2}P(X = Y) \leq \frac{1}{2},$$

or equivalently,

$$E[F_Y(X)] - E[F_X(X)] \geq \frac{1}{2}(P(X = Y) - P(X = X')). \quad \blacksquare$$

Theorem 3.12 generalises the result established in [54, Equation 12] for continuous and independent random variables. For this particular case, Equation (3.5) can be simplified. The reason is that for continuous and independent random variables X, X' and Y the probabilities $P(X = Y)$ and $P(X = X')$ equals zero, and then the second part of Equation (3.5) is simplified.

Corollary 3.13 *Let X and Y be two real-valued independent and continuous random variables with cumulative distribution functions F_X and F_Y , respectively. Then:*

$$X \succeq_{\text{SP}} Y \Leftrightarrow E[F_Y(X)] \geq E[F_X(X)].$$

If we are dealing with discrete and independent real-valued random variables, Equation (3.5) can also be re-written. Before showing how, let us give the following lemma:

Lemma 3.14 *Let $\{p_n\}_{n \in \mathbb{N}}$ be a sequence of positive real numbers such that $\sum_n p_n = 1$. Then it holds that:*

$$1 = \sum_n p_n^2 + 2 \sum_{n < m} p_n p_m.$$

Proof: The result is a direct consequence of:

$$1 = \sum_n p_n = \left(\sum_n p_n \right)^2 = \sum_n p_n^2 + 2 \sum_{n < m} p_n p_m. \quad \blacksquare$$

Proposition 3.15 *Let X and Y be two real-valued discrete and independent random variables. If S_X denotes the support of X , then $X \succeq_{\text{SP}} Y$ holds if and only if*

$$E[F_Y(X^-) - F_X(X^-)] \geq \frac{1}{2} \sum_{x \in S_X} P(X = x)(P(Y = x) - P(X = x)),$$

where $F_X(t^-)$ and $F_Y(t^-)$ denote the left hand side limit of the cumulative distribution functions F_X and F_Y evaluated in t . That is:

$$F_X(t^-) = P(X < t) \text{ and } F_Y(t^-) = P(Y < t).$$

Proof: Applying the definition of the probabilistic relation Q :

$$\begin{aligned} Q(X, Y) &= P(X > Y) + \frac{1}{2}P(X = Y) \\ &= \sum_{x \in S_X} P(X = x)P(Y < x) + \frac{1}{2} \sum_{x \in S_X} P(X = x)P(Y = x) \\ &= \sum_{x \in S_X} P(X = x)F_Y(x^-) + \frac{1}{2} \sum_{x \in S_X} P(X = x)P(Y = x). \end{aligned}$$

Thus, $Q(X, Y) \geq \frac{1}{2}$ if and only if:

$$\sum_{x \in S_X} P(X = x)F_Y(x^-) \geq \frac{1}{2} \left(1 - \sum_{x \in S_X} P(X = x)P(Y = x) \right).$$

Applying Lemma 3.14, the right hand side of the previous inequality becomes:

$$\begin{aligned} &\frac{1}{2} \left(\sum_{x \in S_X} P(X = x)^2 + 2 \sum_{x_1, x_2 \in S_X, x_1 < x_2} P(X = x_1)P(X = x_2) \right. \\ &\quad \left. - \sum_{x \in S_X} P(X = x)P(Y = x) \right) = \frac{1}{2} \left(\sum_{x \in S_X} P(X = x)^2 \right. \\ &\quad \left. + 2 \sum_{x \in S_X} P(X = x)F_X(x^-) - \sum_{x \in S_X} P(X = x)P(Y = x) \right) \\ &= E[F_X(X^-)] + \frac{1}{2} \left(\sum_{x \in S_X} P(X = x)(P(X = x) - P(Y = x)) \right). \end{aligned}$$

Then, it holds that $Q(X, Y) \geq \frac{1}{2}$ if and only if

$$E[F_Y(X^-) - F_X(X^-)] \geq \frac{1}{2} \sum_{x \in S_X} P(X = x)(P(X = x) - P(Y = x)). \quad \blacksquare$$

Theorem 3.12 allows to characterise statistical preference between independent random variables. However, we have already said that statistical preference is a method that considers the joint distribution of the random variables. For this reason, we are interested not only in independent random variables but also in dependent ones. Next, we focus on comonotonic and countermonotonic random variables, that correspond to the extreme cases of joint distribution functions according to the Fréchet-Hoeffding bounds given in Equation (2.8).

Continuous comonotonic and countermonotonic random variables

Let us consider two continuous random variables whose cumulative distribution functions are F_X and F_Y , respectively, and f_X and f_Y denote their respective density functions.

First of all, let us consider the case in which X and Y are comonotonic. Then, the joint cumulative distribution function of X and Y is:

$$F_{X,Y}(x, y) = \min(F_X(x), F_Y(y)), \text{ for every } x, y \in \mathbb{R}.$$

The value of the relation $Q(X, Y)$ has already been studied by De Meyer et al.

Proposition 3.16 ([54, Prop.7]) *Let X and Y be two real-valued comonotonic and continuous random variables. The probabilistic relation $Q(X, Y)$ has the following expression:*

$$Q(X, Y) = \int_{x:F_X(x) < F_Y(x)} f_X(x) dx + \frac{1}{2} \int_{x:F_X(x) = F_Y(x)} f_X(x) dx. \quad (3.6)$$

In fact, it holds that:

$$\begin{aligned} P(X > Y) &= \int_{x:F_X(x) < F_Y(x)} f_X(x) dx \text{ and} \\ P(X = Y) &= \int_{x:F_X(x) = F_Y(x)} f_X(x) dx. \end{aligned}$$

Therefore, we obtain that $X \succeq_{\text{SP}} Y$ if and only if Equation (3.6) takes a value greater than or equal to $\frac{1}{2}$. However, by Lemma 2.20 we know that $X \succeq_{\text{SP}} Y$ if and only if $Q(X, Y) \geq Q(Y, X)$. These are given by:

$$\begin{aligned} Q(X, Y) &= \int_{x:F_X(x) < F_Y(x)} f_X(x) dx + \frac{1}{2} \int_{x:F_X(x) = F_Y(x)} f_X(x) dx. \\ Q(Y, X) &= \int_{x:F_Y(x) < F_X(x)} f_Y(x) dx + \frac{1}{2} \int_{x:F_Y(x) = F_X(x)} f_Y(x) dx \\ &= 1 - \int_{x:F_X(x) < F_Y(x)} f_Y(x) dx - \frac{1}{2} \int_{x:F_Y(x) = F_X(x)} f_Y(x) dx. \end{aligned}$$

Hence, we obtain the following:

Corollary 3.17 *Let X and Y be two real-valued comonotonic and continuous random variables, where F_X and F_Y denote their respective cumulative distribution functions and f_X and f_Y denote their respective density functions. Then, $X \succeq_{\text{SP}} Y$ if and only if:*

$$\int_{x:F_X(x) < F_Y(x)} (f_X(x) + f_Y(x)) dx + \frac{1}{2} \int_{x:F_X(x) = F_Y(x)} (f_X(x) + f_Y(x)) dx \geq 1.$$

Assume now that X and Y are continuous and countermonotonic real-valued random variables. In that case, the joint cumulative distribution function is given by:

$$F_{X,Y}(x, y) = \max(F_X(x) + F_Y(y) - 1, 0), \text{ for } x, y \in \mathbb{R}.$$

As in the case of comonotonic random variables, De Meyer et al. also found the expression of $Q(X, Y)$.

Proposition 3.18 ([54, Prop.7]) *Let X and Y be two real-valued countermonotonic and continuous random variables. The probabilistic relation $Q(X, Y)$ is given by:*

$$Q(X, Y) = F_Y(u), \quad (3.7)$$

where u is one point that fulfills $F_X(u) + F_Y(u) = 1$.

Therefore, using Equation (3.7) it is possible to state the following proposition.

Proposition 3.19 *Let X and Y be two real-valued countermonotonic and continuous random variables. If F_X and F_Y denote their respective cumulative distribution functions, the following equivalence holds:*

$$X \succeq_{\text{SP}} Y \Leftrightarrow F_Y(u) \geq F_X(u),$$

where u is a point such that $F_X(u) + F_Y(u) = 1$.

Proof: By definition, $X \succeq_{\text{SP}} Y$ if and only if $Q(X, Y) \geq \frac{1}{2}$. However, using Equation (3.7), $Q(X, Y) \geq \frac{1}{2}$ is equivalent to $F_Y(u) \geq \frac{1}{2}$. But, since u satisfies $F_X(u) + F_Y(u) = 1$, $F_Y(u) \geq \frac{1}{2}$ if and only if $F_Y(u) \geq F_X(u)$. ■

Discrete comonotonic and countermonotonic random variables with finite supports

In the previous paragraph we considered continuous comonotonic and countermonotonic random variables, and we characterised statistical preference for them. Now, we also consider real-valued random variables coupled by the minimum or Łukasiewicz operators, but we assume them to be discrete with finite supports. For these variables, De Meyer et al. also found the expression of the probabilistic relation Q .

Proposition 3.20 ([54, Prop. 2]) *Let X and Y be two real-valued comonotonic and discrete random variables with finite supports. Then, their supports, denoted by S_X and S_Y , respectively, can be expressed by:*

$$S_X = \{x_1, \dots, x_n\} \text{ and } S_Y = \{y_1, \dots, y_n\}$$

such that $x_1 \leq \dots \leq x_n$ and $y_1 \leq \dots \leq y_n$, and such that

$$P(X = x_i) = P(Y = y_i) = P(X = x_i, Y = y_i), \text{ for } i = 1, \dots, n.$$

Furthermore, the probabilistic relation takes the value:

$$Q(X, Y) = \sum_{i=1}^n P(X = x_i) \delta_i^M, \quad (3.8)$$

where

$$\delta_M^i = \begin{cases} 1 & \text{if } x_i > y_i. \\ \frac{1}{2} & \text{if } x_i = y_i. \\ 0 & \text{if } x_i < y_i. \end{cases}$$

The following example illustrates this result.

Example 3.21 ([54, Example 3]) Consider the comonotonic random variables X and Y defined by:

X	1	3	4	Y	2	3	5
P_X	0.15	0.4	0.45	P_Y	0.35	0.35	0.3

De Schuymer et al. proved that their supports, S_X and S_Y , respectively, can be expressed by:

$$S_X = \{x_1, x_2, x_3, x_4, x_5\} = \{1, 3, 3, 4, 4\} \text{ and } S_Y = \{y_1, y_2, y_3, y_4, y_5\} = \{2, 2, 3, 3, 5\},$$

and their probabilities can be expressed by:

X	x_1	x_2	x_3	x_4	x_5	Y	y_1	y_2	y_3	y_4	y_5
P_X	0.15	0.2	0.2	0.15	0.3	P_Y	0.15	0.2	0.2	0.15	0.3

Using the notation of the previous result, it holds that:

$$\begin{aligned} \delta_1^M &= 0 \text{ because } x_1 < y_1, & \delta_4^M &= 1 \text{ because } x_4 > y_4. \\ \delta_2^M &= 1 \text{ because } x_2 > y_2, & \delta_5^M &= 0 \text{ because } x_5 < y_5. \\ \delta_3^M &= 0.5 \text{ because } x_3 = y_3. \end{aligned}$$

Then:

$$\begin{aligned} Q(X, Y) &= \sum_{i=1}^5 \delta_i^M P(X = x_i) = P(X = x_2) + \frac{1}{2}P(X = x_3) + P(X = x_4) \\ &= 0.2 + \frac{1}{2}0.2 + 0.15 = 0.45. \blacklozenge \end{aligned}$$

Under the previous conditions, it is possible to define the probability space $(\Omega^*, \mathcal{P}(\Omega^*), P_1)$, where $\Omega^* = \{\omega_1, \dots, \omega_n\}$ and

$$P_1(\{\omega_i\}) = P(X = x_i), \text{ for any } i = 1, \dots, n.$$

We can also define the random variables X^* and Y^* by:

$$X^*(\omega_i) = x_i \text{ and } Y^*(\omega_i) = y_i \text{ for any } i = 1, \dots, n.$$

Then, the random variables X^* and Y^* are equally distributed than X and Y , respectively. This will be a very important fact for results in Section 3.2. Next lemma proves that $Q(X, Y) = Q(X^*, Y^*)$.

Lemma 3.22 *Under the previous conditions, it holds that $Q(X, Y) = Q(X^*, Y^*)$.*

Proof: Let us compute the value of $P^*(X^* > Y^*)$ and $P^*(X^* = Y^*)$:

$$\begin{aligned} P_1(X^* > Y^*) &= P_1(\{\omega_i : X^*(\omega_i) = x_i > y_i = Y^*(\omega_i)\}) \\ &= \sum_{i=1}^n P_1(\{\omega_i\}) I_{x_i > y_i} = \sum_{i=1}^n P(X = x_i) I_{x_i > y_i}. \\ P_1(X^* = Y^*) &= P_1(\{\omega_i : X^*(\omega_i) = x_i = y_i = Y^*(\omega_i)\}) \\ &= \sum_{i=1}^n P_1(\{\omega_i\}) I_{x_i = y_i} = \sum_{i=1}^n P(X = x_i) I_{x_i = y_i}. \end{aligned}$$

Then:

$$\begin{aligned} Q(X^*, Y^*) &= P^*(X^* > Y^*) + \frac{1}{2} P^*(X^* = Y^*) \\ &= \sum_{i=1}^n P(X = x_i) I_{x_i > y_i} + \frac{1}{2} \sum_{i=1}^n P(X = x_i) I_{x_i = y_i} \\ &= \sum_{i=1}^n P(X = x_i) \delta_i^M = Q(X, Y). \quad \blacksquare \end{aligned}$$

Example 3.23 *Let us continue with Example 3.21. We have two random variables X and Y and we have seen that their supports can be expressed by $S_X = \{x_1, \dots, x_5\} = \{1, 3, 3, 4, 4\}$ and $S_Y = \{y_1, \dots, y_5\} = \{2, 2, 3, 3, 5\}$, respectively. Their probability distributions are given by:*

X	x_1	x_2	x_3	x_4	x_5	Y	y_1	y_2	y_3	y_4	y_5
P_X	0.15	0.2	0.2	0.15	0.3	P_Y	0.15	0.2	0.2	0.15	0.3

Now, we can define the possibility space $\Omega^* = \{\omega_1, \dots, \omega_5\}$, the probability P_1 such that $P_1(\omega_i) = P(X = x_i)$ and the random variables X^* and Y^* by:

$$X^*(\omega_i) = x_i \text{ and } Y^*(\omega_i) = y_i \text{ for any } i = 1, \dots, 5.$$

Now, taking into account that:

$$\begin{aligned} x_1 = 1 < 2 = y_1 &\Rightarrow \delta_1^M = 0 \text{ and } X^*(\omega_1) < Y^*(\omega_1), \\ x_2 = 3 > 2 = y_2 &\Rightarrow \delta_2^M = 1 \text{ and } X^*(\omega_2) > Y^*(\omega_2), \\ x_3 = 3 = y_3 &\Rightarrow \delta_3^M = \frac{1}{2} \text{ and } X^*(\omega_3) = Y^*(\omega_3), \\ x_4 = 4 > 3 = y_4 &\Rightarrow \delta_4^M = 1 \text{ and } X^*(\omega_4) > Y^*(\omega_4), \\ x_5 = 4 < 5 = y_5 &\Rightarrow \delta_5^M = 0 \text{ and } X^*(\omega_5) < Y^*(\omega_5), \end{aligned}$$

it is possible to compute the value of the probabilistic relation $Q(X^*, Y^*)$:

$$\begin{aligned} Q(X^*, Y^*) &= P_1(X^* > Y^*) + \frac{1}{2} P_1(X^* = Y^*) \\ &= P_1(\{\omega_2, \omega_4\}) + \frac{1}{2} P_1(\{\omega_3\}) = 0.2 + 0.15 + \frac{1}{2} 0.2 = 0.45 < \frac{1}{2}, \end{aligned}$$

hence $Y^* \succ_{\text{SP}} Y^*$. Furthermore, in Example 3.21 we obtained that $Q(X, Y) = 0.45$, and therefore, by the previous lemma, it holds that $Q(X, Y) = Q(X^*, Y^*) = 0.45$. \blacklozenge

Remark 3.24 Taking the previous comments into account, we shall assume without loss of generality that any two discrete and comonotonic random variables X and Y with finite supports are defined in a probability space $(\Omega, \mathcal{P}(\Omega), P)$, where Ω is finite, $\Omega = \{\omega_1, \dots, \omega_n\}$, and $X(\omega_i) = x_i$, $Y(\omega_i) = y_i$, such that $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for any $i = 1, \dots, n-1$. Moreover:

$$P(X = x_i, Y = y_i) = P(X = x_i) = P(Y = y_i) \text{ for } i = 1, \dots, n.$$

Furthermore, $Q(X, Y)$ is given by Equation (3.8). \blacklozenge

Next result gives a characterization of statistical preference in terms of the supports of X and Y , and also in terms of the probability measure in the initial space. Its proof is trivial and therefore omitted.

Proposition 3.25 Consider two real-valued comonotonic and discrete random variables X and Y with finite supports. According to the previous remark, we can assume them to be defined on $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \dots, \omega_n\}$, by $X(\omega_i) = x_i$ and $Y(\omega_i) = y_i$, where $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for any $i = 1, \dots, n-1$. Then, $X \succeq_{\text{SP}} Y$ if and only if:

$$\sum_{i: x_i > y_i} P(X = x_i) \geq \sum_{i: x_i < y_i} P(X = x_i),$$

or equivalently, by Lemma 3.22, if and only if:

$$\sum_{i: x_i > y_i} P(\{\omega_i\}) \geq \sum_{i: x_i < y_i} P(\{\omega_i\}).$$

Now, we focus on countermonotonic random variables. For them, De Meyer et al. proved the following result:

Proposition 3.26 ([54, Prop. 4]) Let X and Y be real-valued comonotonic and discrete random variables with finite supports. Then, their supports can be expressed by $S_X = \{x_1, \dots, x_n\}$ and $S_Y = \{y_1, \dots, y_n\}$, respectively, such that $x_1 \leq \dots \leq x_n$ and $y_1 \leq \dots \leq y_n$, and such that:

$$P(X = x_i) = P(Y = y_{n-i+1}) = P(X = x_i, Y = y_i)$$

for any $i = 1, \dots, n$. Under these conditions, the probabilistic relation $Q(X, Y)$ takes the value:

$$Q(X, Y) = \sum_{i=1}^n P(X = x_i) \delta_i^L, \quad (3.9)$$

where

$$\delta_i^L = \begin{cases} 1 & \text{if } x_i > y_{n-i+1}. \\ \frac{1}{2} & \text{if } x_i = y_{n-i+1}. \\ 0 & \text{if } x_i < y_{n-i+1}. \end{cases}$$

To illustrate this result, consider the following example.

Example 3.27 ([54, Example 5]) Consider the random variables X and Y of Example 3.21, but now assume them to be countermonotonic. Their supports can be expressed by $S_X = \{x_1, x_2, x_3, x_4, x_5\} = \{1, 3, 3, 4, 4\}$ and $S_Y = \{y_1, y_2, y_3, y_4, y_5\} = \{2, 3, 3, 5, 5\}$. Furthermore, the probability distributions of X and Y can be expressed by:

$$\begin{array}{c|ccccc} X & x_1 & x_2 & x_3 & x_4 & x_5 \\ \hline P_X & 0.15 & 0.15 & 0.25 & 0.1 & 0.35 \end{array} \quad \begin{array}{c|ccccc} Y & y_1 & y_2 & y_3 & y_4 & y_5 \\ \hline P_Y & 0.35 & 0.1 & 0.25 & 0.15 & 0.15 \end{array}$$

Using the notation of the previous result, it holds that:

$$\begin{aligned} \delta_1^L &= 0 \text{ because } x_1 < y_5, & \delta_4^L &= 1 \text{ because } x_4 > y_4. \\ \delta_2^L &= 0 \text{ because } x_2 < y_4, & \delta_5^L &= 1 \text{ because } x_5 > y_5. \\ \delta_3^L &= 0.5 \text{ because } x_3 = y_3. \end{aligned}$$

Then:

$$\begin{aligned} Q(X, Y) &= \sum_{i=1}^5 \delta_i^L P(X = x_i) = \frac{1}{2} P(X = x_3) + P(X = x_4) + P(X = x_5) \\ &= \frac{1}{2} 0.25 + 0.1 + 0.35 = 0.575. \blacklozenge \end{aligned}$$

Under the above conditions, and similarly to the case of comonotonic random variables, it is possible to define a probability space $(\Omega^*, \mathcal{P}(\Omega^*), P_2)$, where $\Omega^* = \{\omega_1, \dots, \omega_n\}$ and the probability is given by:

$$P_2(\{\omega_i\}) = P(X = x_i) \text{ for every } i = 1, \dots, n.$$

Furthermore, we can also define the random variables X^* and Y^* by:

$$X^*(\omega_i) = x_i \text{ and } Y^*(\omega_i) = y_{n-i+1} \text{ for any } i = 1, \dots, n.$$

Note that the variables X and X^* , and also Y and Y^* , are equally distributed. Furthermore, next lemma shows that $Q(X, Y) = Q(X^*, Y^*)$.

Lemma 3.28 In the conditions of the previous comments, considering the probability space $(\Omega^*, \mathcal{P}(\Omega^*), P_2)$ and the random variables X^* and Y^* , it holds that $Q(X, Y) = Q(X^*, Y^*)$.

Proof: Let us compute the value of $P_2(X^* > Y^*)$ and $P_2(X^* = Y^*)$:

$$\begin{aligned} P_2(X^* > Y^*) &= P_2(\{\omega_i : X^*(\omega_i) = x_i > y_{n-i+1} = Y^*(\omega_i)\}) \\ &= \sum_{i=1}^n P_2(\{\omega_i\}) I_{x_i > y_{n-i+1}} = \sum_{i=1}^n P(X = x_i) I_{x_i > y_{n-i+1}}. \\ P_2(X^* = Y^*) &= P_2(\{\omega_i : X^*(\omega_i) = x_i = y_{n-i+1} = Y^*(\omega_i)\}) \\ &= \sum_{i=1}^n P_2(\{\omega_i\}) I_{x_i = y_{n-i+1}} = \sum_{i=1}^n P(X = x_i) I_{x_i = y_{n-i+1}}. \end{aligned}$$

Then:

$$\begin{aligned} Q(X^*, Y^*) &= P_2(X^* > Y^*) + \frac{1}{2} P_2(X^* = Y^*) \\ &= \sum_{i=1}^n P(X = x_i) I_{x_i > y_{n-i+1}} + \frac{1}{2} \sum_{i=1}^n P(X = x_i) I_{x_i = y_{n-i+1}} \\ &= \sum_{i=1}^n P(X = x_i) \delta_i^L = Q(X, Y). \quad \blacksquare \end{aligned}$$

Next example helps to understand how to build the probability space and the random variables.

Example 3.29 Consider again Example 3.27. The supports of the random variables X and Y can be expressed by $S_X = \{x_1, \dots, x_5\} = \{1, 3, 3, 4, 4\}$ and $S_Y = \{y_1, \dots, y_5\} = \{2, 3, 3, 5, 5\}$, respectively. Their probability distributions are given by:

X	x_1	x_2	x_3	x_4	x_5	Y	y_1	y_2	y_3	y_4	y_5
P_X	0.15	0.15	0.25	0.1	0.35	P_Y	0.35	0.1	0.25	0.15	0.15

Now, we can define the possibility space $\Omega^* = \{\omega_1, \dots, \omega_5\}$, the probability P^* satisfying that $P^*(\{\omega_i\}) = P(X = x_i)$ for $i = 1, \dots, 5$ and the random variables X^* and Y^* by:

$$X^*(\omega_i) = x_i \text{ and } Y^*(\omega_i) = y_{6-i} \text{ for any } i = 1, \dots, 5.$$

Taking into account that:

$$\begin{aligned} x_1 = 1 < 5 = y_5 &\Rightarrow \delta_1^L = 0 \text{ and } X^*(\omega_1) < Y^*(\omega_1), \\ x_2 = 3 < 5 = y_4 &\Rightarrow \delta_2^L = 0 \text{ and } X^*(\omega_2) < Y^*(\omega_2), \\ x_3 = 3 = y_3 &\Rightarrow \delta_3^L = \frac{1}{2} \text{ and } X^*(\omega_3) = Y^*(\omega_3), \\ x_4 = 4 > 3 = y_2 &\Rightarrow \delta_4^L = 1 \text{ and } X^*(\omega_4) > Y^*(\omega_4), \\ x_5 = 4 > 2 = y_1 &\Rightarrow \delta_5^L = 1 \text{ and } X^*(\omega_5) > Y^*(\omega_5), \end{aligned}$$

it is possible to compute the value of the probabilistic relation $Q(X^*, Y^*)$:

$$\begin{aligned} Q(X^*, Y^*) &= P^*(X^* > Y^*) + \frac{1}{2}P^*(X^* = Y^*) \\ &= \frac{1}{2}P^*(\{\omega_3\}) + P^*(\{\omega_4, \omega_5\}) = \frac{1}{2}0.25 + 0.1 + 0.35 = 0.575 > \frac{1}{2}, \end{aligned}$$

whence $Y^* \succ_{\text{SP}} Y^*$. Moreover, from Example 3.27 $Q(X, Y) = 0.575$, and therefore, as we have seen in the previous lemma, $Q(X, Y) = Q(X^*, Y^*) = 0.575$. \blacklozenge

Remark 3.30 Using the previous result we can assume, without loss of generality, that any two countermonotonic real-valued random variables X and Y are defined on a probability space $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \dots, \omega_n\}$, by $X(\omega_i) = x_i$ and $Y(\omega_i) = y_{n-i+1}$ such that $x_i \leq x_{i+1}$ and $y_i \geq y_{i+1}$ for $i = 1, \dots, n$, and satisfying that

$$P(X = x_i, Y = y_i) = P(X = x_i) = P(Y = y_{n-i+1}) \text{ for } i = 1, \dots, n. \blacklozenge$$

Now, assuming the conditions of the previous remark, we prove that there is, at most, one element ω_i such that $X(\omega_i) = Y(\omega_i)$.

Lemma 3.31 In the conditions of the previous remark, if there exists $l > 0$ such that

$$\begin{aligned} X(\omega_k) &= \dots = X(\omega_{k+l}) = Y(\omega_k) = \dots = Y(\omega_{k+l}), \\ \min(|X(\omega_{k-1}) - X(\omega_{k+l+1})|, |Y(\omega_{k-1}) - Y(\omega_{k+l+1})|) &> 0, \end{aligned}$$

for some k , then it is possible to define a probability space $(\Omega^*, \mathcal{P}(\Omega^*), P_3)$ and two random variables X^* and Y^* such that:

- $Q(X^*, Y^*) = Q(X, Y)$.
- There are not $\omega', \omega'' \in \Omega^*$ such that

$$X^*(\omega') = X^*(\omega'') = Y^*(\omega') = Y^*(\omega'').$$

- X^* and Y^* follow the same distribution than X and Y , respectively.

Proof: Define $\Omega^* = \{\omega_1^*, \dots, \omega_{n-l}^*\}$ and let P_3 be the probability given by:

$$\begin{aligned} P_3(\{\omega_i^*\}) &= P(\{\omega_i\}) \text{ for any } i = 1, \dots, k-1. \\ P_3(\{\omega_k^*\}) &= P(\{\omega_k\}) + \dots + P(\{\omega_{k+l}\}). \\ P_3(\{\omega_i^*\}) &= P(\{\omega_{i+1}\}) \text{ for any } i = k+l+1, \dots, n-1. \end{aligned}$$

Consider the random variables X^* and Y^* given by:

$$\begin{aligned} X^*(\omega_i^*) &= X(\omega_i) \text{ and } Y^*(\omega_i^*) = Y(\omega_i) \text{ for any } i = 1, \dots, k-1. \\ X^*(\omega_k^*) &= X(\omega_k) \text{ and } Y^*(\omega_k^*) = Y(\omega_k). \\ X^*(\omega_i^*) &= X(\omega_{i+1}) \text{ and } Y^*(\omega_i^*) = Y(\omega_{i+1}) \text{ for any } i = k+l+1, \dots, n-1. \end{aligned}$$

They satisfy that:

$$\begin{aligned} X^*(\omega_i^*) &< Y^*(\omega_i^*) \text{ for any } i = 1, \dots, k-1. \\ X^*(\omega_k^*) &= Y^*(\omega_k^*). \\ X^*(\omega_i^*) &> Y^*(\omega_i^*) \text{ for any } i = k+l+1, \dots, n-1. \end{aligned}$$

Then, since

$$\omega_k^* \notin \{\omega^* \in \Omega^* : X^*(\omega^*) > Y^*(\omega^*)\} \text{ and} \\ \omega_k, \dots, \omega_{k+l} \notin \{\omega \in \Omega : X(\omega) > Y(\omega)\},$$

it holds that:

$$\omega_i^* \in \{X^* > Y^*\} \Leftrightarrow \omega_i \in \{X > Y\}, \text{ for } i = 1, \dots, k-1.$$

Furthermore, $\omega_{i-1}^* \notin \{X^* > Y^*\}$ and $\omega_i \notin \{X > Y\}$ for $i = 1, \dots, k-1$. Then, we conclude that:

$$\begin{aligned} P_3(X^* > Y^*) &= P_3(\{\omega^* \in \Omega^* : X^*(\omega^*) > Y^*(\omega^*)\}) \\ &= \sum_{i: X^*(\omega_i^*) > Y^*(\omega_i^*)} P_3(\{\omega_i^*\}) = \sum_{i: X(\omega_i) > Y(\omega_i)} P(\{\omega_i\}) = P(X > Y). \end{aligned}$$

Furthermore, since $X^*(\omega_k^*) = Y^*(\omega_k^*)$ and $P_3(\{\omega_k^*\}) = P(\{\omega_k, \dots, \omega_{k+l}\})$, it holds that:

$$\begin{aligned} P_3(X^* = Y^*) &= P_3(\{\omega^* \in \Omega^* : X^*(\omega^*) = Y^*(\omega^*)\}) \\ &= \sum_{i: X^*(\omega_i^*) = Y^*(\omega_i^*)} P_3(\{\omega_i^*\}) = P_3(\{\omega_k^*\}) \\ &= P(\{\omega_k\}) + \dots + P(\{\omega_{k+l}\}) = \sum_{i: X'(\omega_i) = Y'(\omega_i)} P(\{\omega_i\}) = P(X' = Y'). \end{aligned}$$

Then, $Q(X^*, Y^*) = Q(X, Y)$.

Moreover, by construction there are not $\omega', \omega'' \in \Omega^*$, $\omega' \neq \omega''$, such that

$$X^*(\omega') = X^*(\omega'') = Y^*(\omega') = Y^*(\omega'').$$

Finally, it is obvious that X^* and X , and also Y^* and Y , are equally distributed, since they take the same values with the same probabilities. ■

Remark 3.32 Taking into account the previous result and Remark 3.30, we conclude that given two discrete countermonotonic random variables X and Y with finite supports, we can assume, without loss of generality, that their supports are given by $S_X = \{x_1, \dots, x_n\}$ and $S_Y = \{y_1, \dots, y_n\}$, where $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for $i = 1, \dots, n-1$, and that they are defined in a probability space $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \dots, \omega_n\}$, by $X(\omega_i) = x_i$ and $Y(\omega_i) = y_{n-i+1}$. Furthermore:

$$P(X = x_i, Y = y_i) = P(X = x_i) = P(Y = y_{n-i+1}) \text{ for any } i = 1, \dots, n.$$

Under these conditions, $Q(X, Y)$ is given by Equation (3.9). Furthermore, using the previous lemma we can also assume that $\max\{|X(\omega_i) - X(\omega_{i+1})|, |Y(\omega_i) - Y(\omega_{i+1})|\} > 0$ for any $i = 1, \dots, n-1$. ♦

These results allow us to characterise statistical preference for discrete countermonotonic random variables with finite supports.

Proposition 3.33 *Let X and Y be two real-valued discrete and countermonotonic random variables with finite supports, that can be expressed as in the previous remark. Then, it is possible to characterise $X \succeq_{\text{SP}} Y$ in the following way:*

- *If there exists k such that $X(\omega_k) = Y(\omega_k)$, then $X \succeq_{\text{SP}} Y$ if and only if:*

$$P(X = x_1) + \dots + P(X = x_{k-1}) \leq P(X = x_{k+1}) + \dots + P(X = x_n),$$

or equivalently, if and only if:

$$P(\{\omega_1\}) + \dots + P(\{\omega_{k-1}\}) \leq P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\}).$$

- *If $X(\omega_i) \neq Y(\omega_i)$ for any $i = 1, \dots, n$, denote by $k = \min\{i : X(\omega_i) < Y(\omega_i)\}$. Then $X \succeq_{\text{SP}} Y$ if and only if:*

$$P(X = x_1) + \dots + P(X = x_k) \leq P(X = x_{k+1}) + \dots + P(X = x_n),$$

or equivalently, if and only if:

$$P(\{\omega_1\}) + \dots + P(\{\omega_k\}) \leq P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\}).$$

Proof: Assume that there is k such that $X(\omega_k) = Y(\omega_k)$. Then, $X(\omega_i) > Y(\omega_i)$ for any $i < k$ and $X(\omega_i) < Y(\omega_i)$ for any $i > k$. Then:

$$\begin{aligned} Q(X, Y) &= P(\{\omega_{k+1}, \dots, \omega_n\}) + \frac{1}{2}P(\{\omega_k\}) \text{ and} \\ Q(Y, X) &= P(\{\omega_1, \dots, \omega_{k-1}\}) + \frac{1}{2}P(\{\omega_k\}). \end{aligned}$$

Then, $Q(X, Y) \geq \frac{1}{2}$ if and only if:

$$P(\{\omega_{k+1}, \dots, \omega_n\}) \geq P(\{\omega_1, \dots, \omega_{k-1}\}).$$

Furthermore, the previous expression is equivalent to:

$$P(X = x_{k+1}) + \dots + P(X = x_n) \geq P(X = x_1) + \dots + P(X = x_{k-1}).$$

Now, assume that $X(\omega_i) \neq Y(\omega_i)$ for any $i = 1, \dots, n$. Then, denote by k the element $k = \max\{i : X(\omega_i) < Y(\omega_i)\}$. Then, $X(\omega_i) > Y(\omega_i)$ for any $i = k + 1, \dots, n$ and $X(\omega_i) < Y(\omega_i)$ for any $i = 1, \dots, k$. Then:

$$Q(X, Y) = P(\{\omega_{k+1}, \dots, \omega_n\}) \text{ and } Q(Y, X) = P(\{\omega_1, \dots, \omega_k\}).$$

Then, $Q(X, Y) \geq \frac{1}{2}$ if and only if:

$$P(\{\omega_{k+1}, \dots, \omega_n\}) \geq P(\{\omega_1, \dots, \omega_k\}).$$

This expression is equivalent to:

$$P(X = x_{k+1}) + \dots + P(X = x_n) \geq P(X = x_1) + \dots + P(X = x_k). \quad \blacksquare$$

Random variables coupled by a strict Archimedean copula

Consider two continuous real-valued random variables X and Y with cumulative distribution functions F_X and F_Y , respectively. Let us denote their density functions by f_X and f_Y , respectively. We shall assume the existence of a strict Archimedean copula C , generated by the twice differentiable generator φ , such that

$$F_{X,Y}(x, y) = \varphi^{-1}(\varphi(F_X(x)) + \varphi(F_Y(y))), \text{ for every } x, y \in \mathbb{R}.$$

Note that since C is strict, then $\varphi(0) = \infty$. In that case, we have already mentioned in Equation (2.10) that the pseudo-inverse becomes the inverse, and then the joint cumulative distribution function is given by:

$$F_{X,Y}(x, y) = \varphi^{-1}(\varphi(F_X(x)) + \varphi(F_Y(y))), \text{ for every } x, y \in \mathbb{R}.$$

Now, we are going to obtain the joint density function for (X, Y) . For this aim, we derive $F_{X,Y}$ with respect to x and y :

$$\begin{aligned} \frac{\partial F_{X,Y}}{\partial x}(x, y) &= \frac{\partial \varphi^{-1}(\varphi(F_X(x)) + \varphi(F_Y(y)))}{\partial x}(x, y) \\ &= (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(y)))\varphi'(F_X(x))f_X(x). \\ \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) &= (\varphi^{-1})''(\varphi(F_X(x)) + \varphi(F_Y(y)))\varphi'(F_X(x))\varphi'(F_Y(y))f_X(x)f_Y(y). \end{aligned}$$

Then, the function $f_{X,Y}$ defined by:

$$f_{X,Y}(x, y) = (\varphi^{-1})''(\varphi(F_X(x)) + \varphi(F_Y(y)))\varphi'(F_X(x))\varphi'(F_Y(y))f_X(x)f_Y(y), \quad (3.10)$$

is a density function of (X, Y) . Let us check that $f_{X,Y}(x, y) \geq 0$ for every $x, y \in \mathbb{R}$:

- $f_X, f_Y \geq 0$ because they are density functions.
- By Definition 2.26, $-\varphi$ is 2-monotone. Then, $(-1)^2(-\varphi)' = -\varphi' \geq 0$, that implies $\varphi \leq 0$. Then, $\varphi'(F_X(x))\varphi'(F_Y(y)) \geq 0$.

- Since $-\varphi$ is 2-monotone, $(-1)^3(-\varphi)'' \geq 0$, and then $\varphi'' \geq 0$. Also, it is known that, for a function g , $g^{-1}'(x) = g'(g^{-1}(x))^{-1}$. Then:

$$(\varphi^{-1})'(x) = \frac{1}{\varphi'(\varphi^{-1}(x))},$$

and since $\varphi' \leq 0$, it holds that $(\varphi^{-1})'(x) \leq 0$. Then:

$$(\varphi^{-1})''(x) = -\frac{\varphi''(\varphi^{-1}(x))(\varphi^{-1})'}{\varphi'(\varphi^{-1}(x))}.$$

The denominator is positive because it is squared. Furthermore, φ'' is positive, but $(\varphi^{-1})'$ is negative, but when multiplying for (-1) it becomes positive.

Then, f is the product of positive elements, and therefore f is positive. Now, let us see that the area below $f_{X,Y}$ is 1:

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x,y) dy dx &= \int_{\mathbb{R}} \left((\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(y))) \right)_{-\infty}^{\infty} \varphi'(F_X(x)) f_X(x) dx \\ &= \int_{\mathbb{R}} (\varphi^{-1})'(\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx \\ &= \varphi^{-1}(\varphi(F_X(x))) \Big|_{-\infty}^{\infty} = F_X(x) \Big|_{-\infty}^{\infty} = 1. \end{aligned}$$

Using the expression of the joint density function in Equation (3.10) we can prove the following characterization of the statistical preference.

Theorem 3.34 *Let X and Y be two real-valued continuous random variables, and let F_X and F_Y denote their respective cumulative distribution functions, and f_X and f_Y are their respective density functions. If they are coupled by a strict Archimedean copula C generated by the twice differentiable function φ , then $X \succeq_{SP} Y$ if and only if:*

$$E \left[\left((\varphi^{-1})'(\varphi(F_X(X)) + \varphi(F_Y(X))) - (\varphi^{-1})'(2\varphi(F_X(X))) \right) \varphi'(F_X(X)) \right] \geq 0. \quad (3.11)$$

Proof: First of all, note that (X, Y) is a continuous random vector with density function $f_{X,Y}$. Then, $P(X = Y) = 0$, and therefore $Q(X, Y) = P(X > Y)$ and $Q(Y, X) = P(Y > X)$.

Denote by A the set $A = \{(x, y) \mid x > y\}$. Then,

$$P(X > Y) = \int_A f_{X,Y}(x, y) dy dx.$$

Thus,

$$\begin{aligned}
P(X > Y) &= \int_A f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^x f_{X,Y}(x, y) dy dx \\
&= \int_{-\infty}^{\infty} \left(\varphi^{-1}(\varphi(F_X(x)) + \varphi(F_Y(y))) \Big|_{-\infty}^x \right) \varphi'(F_X(x)) f_X(x) dx \\
&= \int_{-\infty}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx.
\end{aligned}$$

Furthermore, it holds that

$$\begin{aligned}
\int_{-\infty}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx \\
= \frac{1}{2} \varphi^{-1}(2\varphi(F_X(X))) \Big|_{-\infty}^{\infty} = \frac{1}{2} (\varphi^{-1}(0) - \varphi^{-1}(\infty)) = \frac{1}{2}.
\end{aligned}$$

Therefore, $Q(X, Y) = P(X > Y) \geq \frac{1}{2}$ if and only if

$$\begin{aligned}
&E \left[(\varphi^{-1})'(\varphi(F_X(X)) + \varphi(F_Y(X))) \varphi'(F_X(X)) \right] \\
&= \int_{-\infty}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx \\
&\geq \frac{1}{2} = \int_{-\infty}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx \\
&= E \left[(\varphi^{-1})'(2\varphi(F_X(X))) \varphi'(F_X(X)) \right].
\end{aligned}$$

Hence, this inequality is equivalent to

$$E \left[\left((\varphi^{-1})'(\varphi(F_X(X)) + \varphi(F_Y(X))) - (\varphi^{-1})'(2\varphi(F_X(X))) \right) \varphi'(F_X(X)) \right] \geq 0. \quad \blacksquare$$

This result holds in particular when the random variables are independent, that is, when the copula that links the variables is the product. We have seen in Section 2.1.2 that the product is a strict Archimedean copula with generator $\varphi(t) = -\log t$. In this case:

$$\varphi'(t) = \frac{-1}{t}, \quad \varphi^{-1}(t) = e^{-t} \quad \text{and} \quad (\varphi^{-1})' = -e^t.$$

By replacing these values in Equation (3.11), we obtain that:

$$\begin{aligned}
&(\varphi^{-1})'(\varphi(F_X(X)) + \varphi(F_Y(X))) - (\varphi^{-1})'(2\varphi(F_X(X))) \\
&= -\exp\{\log F_X(X) + \log F_Y(X)\} + \exp\{2 \log F_X(X)\} \\
&= F_Y(X) F_X(X) - F_X(X)^2.
\end{aligned}$$

Then, Equation (3.11) becomes:

$$E \left[(F_Y(X)F_X(X) - F_X(X)^2) \frac{1}{F_X(X)} \right] = E[F_Y(X) - F_X(X)] \geq 0.$$

Thus, we conclude that for continuous random variables X and Y , $X \succeq_{\text{SP}} Y$ if and only if $E[F_Y(X) - F_X(X)] \geq 0$. This result has already been obtained in Corollary 3.13.

Random variables coupled by a nilpotent Archimedean copula

Let us study now the case where the copula that links the real-valued random variables is a nilpotent Archimedean copula generated by a twice differentiable generator. In such case, as we saw in Equation (2.9) the joint distribution function of X and Y is given by:

$$F_{X,Y}(x, y) = \begin{cases} \varphi^{-1}(\varphi(F_X(x)) + \varphi(F_Y(y))) & \text{if } \varphi(F_X(x)) + \varphi(F_Y(y)) \in [0, \varphi(0)]. \\ 0 & \text{otherwise.} \end{cases}$$

Recall that this function cannot be derived in the points (x, y) such that $\varphi(F_X(x)) + \varphi(F_Y(y)) = \varphi(0)$. However, the value of $\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y)$ can be computed for the points (x, y) fulfilling $\varphi(F_X(x)) + \varphi(F_Y(y)) \in [0, \varphi(0))$. In fact, the value of this function is:

$$\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) = (\varphi^{-1})''(\varphi(F_X(x)) + \varphi(F_Y(y))) \varphi'(F_X(x)) \varphi'(F_Y(y)) f_X(x) f_Y(y).$$

In this way, the function $f_{X,Y}$ defined by:

$$f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) & \text{if } \varphi(F_X(x)) + \varphi(F_Y(y)) \in [0, \varphi(0)), \\ 0 & \text{otherwise,} \end{cases}$$

is a joint density function of X and Y : on the one hand, $f_{X,Y}$ is a positive function:

$$\left. \begin{array}{l} f_X, f_Y \geq 0 \\ \varphi' \leq 0 \Rightarrow \varphi'(F_X(x)) \varphi'(F_Y(y)) \geq 0 \\ (\varphi^{-1})'' \geq 0 \end{array} \right\} \Rightarrow f_{X,Y} \geq 0,$$

since it is the product of positive functions. On the other hand, it holds that

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y) dy dx = 1.$$

In order to prove the last equality, we introduce the following notation:

$$y_x = \inf\{y \mid \varphi(F_X(x)) + \varphi(F_Y(y)) \in [0, \varphi(0))\}, \text{ for every } x \in \mathbb{R}.$$

$$s_X = \inf\{x \mid F_X(x) > 0\}.$$

Therefore,

$$\{(x, y) \mid x > s_X, y > y_x\} = \{(x, y) \mid \varphi(F_X(x)) + \varphi(F_Y(y)) < \varphi(0)\}.$$

This implies that:

$$\begin{aligned} & \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y) dy dx \\ &= \int_{s_X}^{\infty} \int_{y_x}^{\infty} (\varphi^{-1})''(\varphi(F_X(x)) + \varphi(F_Y(y))) \varphi'(F_X(x)) \varphi'(F_Y(y)) f_X(x) f_Y(y) dy dx \\ &= \int_{s_X}^{\infty} \left((\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(y))) \Big|_{y_x}^{\infty} \right) \varphi'(F_X(x)) f_X(x) dx \\ &= \int_{s_X}^{\infty} (\varphi^{-1})'(\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx \\ &= \varphi^{-1}(\varphi(F_X(x))) \Big|_{s_X}^{\infty} = F_X(x) \Big|_{s_X}^{\infty} = 1 - F_X(s_X) = 1. \end{aligned}$$

We conclude that $f_{X,Y}$ is a joint density function of X and Y . Let us introduce the following notation:

$$\bar{x} = \inf\{x \mid y_x < x\}. \quad (3.12)$$

Using the function $f_{X,Y}$ and the previous notation, we can prove the following characterization of the statistical preference for random variables coupled by a nilpotent Archimedean copula.

Theorem 3.35 *Let X and Y be two real-valued continuous random variables coupled by a nilpotent Archimedean copula whose generator φ is twice differentiable and φ'' is not the zero function. $X \succeq_{\text{SP}} Y$ if and only if*

$$\int_{\bar{x}}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx \geq \int_{\bar{x}}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx.$$

Proof: From Theorem 3.34, (X, Y) is a continuous random vector with joint density functions $f_{X,Y}$. Then, $P(X = Y) = 0$, and consequently $Q(X, Y) = P(X > Y)$ and $Q(Y, X) = P(Y > X)$.

Let us compute the value of $Q(X, Y) = P(X > Y)$.

$$\begin{aligned}
P(X > Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^x f_{X,Y}(x, y) dy dx \\
&= \int_{\tilde{x}}^{\infty} \int_{y_x}^x (\varphi^{-1})''(\varphi(F_X(x)) + \varphi(F_Y(y))) \varphi'(F_X(x)) \varphi'(F_Y(y)) f_X(x) f_Y(y) dy dx \\
&= \int_{\tilde{x}}^{\infty} \left((\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(y))) \Big|_{y_x}^x \right) \varphi'(F_X(x)) f_X(x) dx \\
&= \int_{\tilde{x}}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx.
\end{aligned}$$

Furthermore, if we denote by \tilde{x} the point

$$\tilde{x} = \inf\{x \mid 2\varphi(F_X(x)) \leq \varphi(0)\}, \quad (3.13)$$

it holds that:

$$\int_{\tilde{x}}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx = \frac{1}{2} \varphi^{-1}(2\varphi(F_X(x))) \Big|_{\tilde{x}}^{\infty} = \frac{1}{2}.$$

For this reason, as $X \succeq_{\text{SP}} Y$ if and only if $Q(X, Y) \geq \frac{1}{2}$, then $X \succeq_{\text{SP}} Y$ if and only if

$$\begin{aligned}
\int_{\tilde{x}}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx &\geq \\
\frac{1}{2} &= \int_{\tilde{x}}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx. \quad \blacksquare
\end{aligned}$$

Remark 3.36 *The previous remark does not generalise Proposition 3.19, where a characterization of statistical preference for continuous and countermonotonic random variables. The reason is that, although the Łukasiewicz operator is an Archimedean copula, its generator is $\varphi(t) = 1 - t$, and $\varphi''(t) = 0$. Hence, this copula does not satisfy the restriction of the previous theorem, which therefore it is not applicable. \blacklozenge*

Characterization of the statistical preference by means of the median

In this section we shall investigate the relationship between statistical preference and the well-know notion of median of a random variable. First of all let us show an example to clarify the connection.

Example 3.37 *Consider again the random variables of Example 2.3. It is easy to check that $Q(X, Y) = 0.6$, and therefore $X \succeq_{\text{SP}} Y$. The intuition here is that in order to obtain $Q(X, Y) = 0.6$ c must be a value greater than 0 and smaller than 1; however, the exact value of $c \in (0, 0.6)$ is not relevant at all.*

Thus, in the discrete case, statistical preference orders the values of the support of X and Y , and once they are ordered, the exact value of each point does not matter: only its relative position and its probability are important. This idea is similar to that used in the definition of the median. \blacklozenge

The first approach to connect statistical preference and the median is to compare the medians of the variables X and Y . Recall that a point t is a median of the random variable X if:

$$P(X \geq t) \geq 0.5 \text{ and } P(X \leq t) \geq 0.5, \quad (3.14)$$

and we denote by $\text{Me}(X)$ the set of medians of the random variable X .

Following the previous example, we conjecture that if the median of X is greater than the median of Y then X should be statistically preferred to Y , and the converse implication should also hold. However, this property does not hold in general.

Remark 3.38 Let X and Y be two real-valued random variables defined on the same probability space. Then there is not a general relationship between $X \succeq_{\text{SP}} Y$ and the following statements:

1. $\text{me}(X) \geq \text{me}(Y)$, for all $\text{me}(X) \in \text{Me}(X)$ and $\text{me}(Y) \in \text{Me}(Y)$.
2. $\text{me}(X) \leq \text{me}(Y)$, for all $\text{me}(X) \in \text{Me}(X)$ and $\text{me}(Y) \in \text{Me}(Y)$.

It is enough to consider the independent random variables X and Y defined in Table 3.1.

X	-2	0	2	Y	-3	1
P_X	0.4	0.2	0.4	P_Y	0.4	0.6

Table 3.1: Definition of random variables X and Y .

Both X and Y have only one median, and they equal to: $\text{me}(X) = 0 < \text{me}(Y) = 1$, but $X \succ_{\text{SP}} Y$ because $Q(X, Y) = 0.64$.

Since both statistical preference and the comparison of medians are complete relations, the same counterexample allows to show that $\text{me}(X) \geq \text{me}(Y)$ does not guarantee that $X \succeq_{\text{SP}} Y$. Notice that $\text{me}(Y) \geq \text{me}(X)$. However, $Q(Y, X) = 0.36$, so that $Y \not\succeq_{\text{SP}} X$.

In order to prove that $X \succeq_{\text{SP}} Y$ and $\text{me}(X) \leq \text{me}(Y)$ are not related in general, it is enough to define X as the constant random variable on 1 and Y as the constant random variable on 0. In this case it is obvious that X and Y have only one median and $\text{me}(X) > \text{me}(Y)$ and $Q(X, Y) = 1$. \blacklozenge

We see thus that statistical preference cannot be reduced to the comparison of the medians of X, Y . Interestingly, there is a connection between statistical preference and the median of $X - Y$, as we shall prove in Theorem 3.40. Let us present a preliminary result.

Proposition 3.39 *Let X and Y be two real-valued random variables defined on the same probability space. Then*

$$X \succeq_{\text{SP}} Y \Leftrightarrow F_{X-Y}(0) \leq F_{Y-X}(0),$$

where F_{X-Y} (respectively, F_{Y-X}) denotes the cumulative distribution function of the random variable $X - Y$ (respectively, $Y - X$).

Proof: By Lemma 2.20, $X \succeq_{\text{SP}} Y$ if and only if $P(X > Y) \geq P(Y > X)$, but:

$$P(X - Y > 0) \geq P(Y - X > 0) \Leftrightarrow 1 - F_{X-Y}(0) \geq 1 - F_{Y-X}(0) \Leftrightarrow F_{X-Y}(0) \leq F_{Y-X}(0).$$

Then, $X \succeq_{\text{SP}} Y$ and $F_{X-Y}(0) \leq F_{Y-X}(0)$ are equivalent. \blacksquare

Therefore, in order to check statistical preference it suffices to evaluate the cumulative distribution functions of $X - Y$ and $Y - X$ on 0. In particular, if $P(X = Y) = 0$, it suffices to evaluate one of the cumulative distribution functions, F_{X-Y} on 0, since in this case,

$$Q(X, Y) = 1 - F_{X-Y}(0)$$

and $X \succeq_{\text{SP}} Y$ if and only if $F_{X-Y}(0) \leq \frac{1}{2}$. This equivalence holds in particular when the random variables form a continuous random vector.

We next prove the connection between statistical preference and the median of $X - Y$.

Theorem 3.40 *Let X and Y be two real-valued random variables defined on the same probability space.*

1. $\sup \text{Me}(X - Y) > 0 \Rightarrow X \succeq_{\text{SP}} Y \Rightarrow \sup \text{Me}(X - Y) \geq 0$.
2. $X \succ_{\text{SP}} Y \Rightarrow \text{Me}(X - Y) \subseteq [0, \infty)$.
3. *The converse implication does not hold, although*

$$\inf \text{Me}(X - Y) > 0 \Rightarrow X \succ_{\text{SP}} Y.$$

4. *If $P(X = Y) = 0$, then*

$$X \succ_{\text{SP}} Y \Leftrightarrow \inf \text{Me}(X - Y) > 0.$$

But even when $P(X = Y) = 0$, $0 \in \text{Me}(X - Y)$ is not equivalent to $Q(X, Y) = \frac{1}{2}$.

Proof:

1. Assume that $\sup \text{Me}(X - Y) > 0$. Then, there is a median $me(X - Y) > 0$. It holds that:

$$\left. \begin{array}{l} P(X > Y) \geq P(X - Y \geq me(X - Y)) \geq \frac{1}{2} \\ P(X < Y) \leq P(X - Y < me(X - Y)) \leq \frac{1}{2} \end{array} \right\} \Rightarrow Q(X, Y) \geq Q(Y, X),$$

and then $X \succeq_{\text{SP}} Y$. Assume that $X \succeq_{\text{SP}} Y$. Then $P(X \geq Y) \geq P(X \leq Y)$. This implies that $P(X - Y \geq 0) \geq Q(X, Y) \geq \frac{1}{2}$, and therefore there exists a median $me(X - Y) \geq 0$, and therefore $\sup \text{Me}(X - Y) \geq me(X - Y) \geq 0$.

2. By definition, $X \succ_{\text{SP}} Y$ if $Q(X, Y) > \frac{1}{2}$.

Now, assume $me(X - Y) < 0$ for a median of $X - Y$, then:

$$\frac{1}{2} \geq P((X - Y) > me(X - Y)) \geq P((X - Y) \geq 0) \geq P(X > Y) + \frac{1}{2}P(X = Y).$$

A contradiction arises because $Q(X, Y) > \frac{1}{2}$.

3. We first prove the implication. Suppose that $me(X - Y) > 0$ for any $me(X - Y) \in \text{Me}(X - Y)$. In such a case:

$$\frac{1}{2} \geq P((X - Y) < me(X - Y)) \geq P(X - Y \leq 0) = 1 - P(X > Y).$$

Hence, $P(X > Y) \geq \frac{1}{2}$ and then $X \succeq_{\text{SP}} Y$. Now, assume that $Q(X, Y) = \frac{1}{2}$. In that case, $P(X \geq Y) = P(Y \geq X) \geq \frac{1}{2}$, and then:

$$P(X - Y \geq 0) = P(Y - X \geq 0) \geq \frac{1}{2},$$

whence $0 \in \text{Me}(X - Y)$, that contradicts the initial hypothesis.

Next, we give an example where $X - Y$ has only one median and equals 0, and $Q(X, Y) < \frac{1}{2}$. It is enough to consider the random variables X and Y whose joint mass function is defined on Table 3.2.

X/Y	0	1	2
0	0.1	0	0.4
1	0	0.4	0
2	0	0	0.1

Table 3.2: Definition of random variables X and Y .

For these variables it holds that $\text{Me}(X - Y) = \{0\}$ but $Y \succ_{\text{SP}} X$, since

$$Q(X, Y) = \frac{1}{2}P((X, Y) = (0, 0), (1, 1), (2, 2)) = \frac{1}{2}0.6 = 0.3 < \frac{1}{2}.$$

4. Assume that $P(X = Y) = 0$ and let us prove the equivalence. On the one hand, assume that $X \succ_{\text{SP}} Y$. By the second item of this Theorem, we know that every median of $X - Y$ is positive. Assume now that 0 is a median of $X - Y$. Then:

$$\frac{1}{2} \geq P(X - Y > 0) = P(X > Y) = Q(X, Y).$$

Then, $Q(X, Y) \leq \frac{1}{2}$, a contradiction. Assume that, although 0 is not a median of $X - Y$, it is the infimum of the medians. In such a case, there is a point $t^* > 0$ such that any point in $(0, t^*]$ is a median of $X - Y$. Then, for any $0 < \varepsilon < t^*$ it holds that:

$$P(X - Y \geq \varepsilon) \geq \frac{1}{2} \text{ and } P(X - Y \leq \varepsilon) \geq \frac{1}{2}.$$

Then, $P(X - Y \geq 0) \geq P(X - Y \geq \varepsilon) \geq \frac{1}{2}$ and:

$$P(X - Y \leq 0) = F_{X-Y}(0) = \lim_{\varepsilon \rightarrow 0} F_{X-Y}(\varepsilon) = \lim_{\varepsilon \rightarrow 0} P(X - Y \leq \varepsilon) \geq \frac{1}{2}.$$

This means that 0 is also a median, and we have already seen that this is not possible. We conclude that $\inf \text{Me}(X - Y) > 0$.

On the other hand, we have seen in the third item that when $\inf \text{Me}(X - Y) > 0$, $X \succ_{\text{SP}} Y$.

Finally, let us see that if 0 is a median of $X - Y$, even when $P(X = Y) = 0$, this is not equivalent to $Q(X, Y) = \frac{1}{2}$. Consider $\Omega = \{\omega_1, \omega_2\}$, the probability measure given by $P(\{\omega_i\}) = \frac{1}{2}$ for $i = 1, 2$, and the random variables X and Y such that $X(\omega_1) = X(\omega_2) = 0$, $Y(\omega_1) = -1$ and $Y(\omega_2) = 1$. Then, -1 is the only median of $X - Y$, and also -1 is the only median of $Y - X$, but $Q(X, Y) = \frac{1}{2}$ and then $X \equiv_{\text{SP}} Y$. On the other hand, consider the space $\Omega = \{\omega_1, \omega_2\}$, $P(\{\omega_1\}) = \frac{3}{4}$ and the random variables defined by:

	ω_1	ω_2
X	0	1
Y	0	0
$X - Y$	0	1

Then, 0 is a median of $X - Y$; however, $Q(X, Y) = \frac{5}{8}$. ■

This theorem establishes a relationship between statistical preference and the median of the difference of the random variables. The particular case in which $P(X = Y) = 0$ is very useful because in that case statistical preference is characterised by the median. Next, we are going to consider two random variables X and Y , and we are going to show how to modify the variables with the aim of avoiding the case $P(X = Y) > 0$.

Lemma 3.41 *Let X, Y be two real-valued discrete random variables, without points of accumulation on their supports, defined on the same probability space such that $P(X =$*

$Y) > 0$. Assume that their supports S_X and S_Y can be expressed by $S_X = \{x_n\}_n$ and $S_Y = \{y_m\}_m$ such that $x_n \leq x_{n+1}$ and $y_m \leq y_{m+1}$ for any n, m . In this case it is possible to build another random variable \tilde{X} fulfilling:

1. $Q(X, Y) = Q(\tilde{X}, Y)$ and

2. $P(X = Y \mid X = x) = 0 \Rightarrow P(\tilde{X} = x) = P(X = x)$.

Proof: We shall use the following notation:

$$P(X = x_n, Y = y_m) = p_{n,m} \text{ for any } n, m.$$

Since $P(X = Y) > 0$, there exists $x_n \in S_X$ and $y_m \in S_Y$ such that $x_n = y_m$ and $p_{n,m} > 0$. Then, for any (x_n, y_m) in this situation we consider $x_n^{(1)}, x_n^{(2)}$ such that:

$$\max\{x_{n-1}, y_{m-1}\} < x_n^{(1)} < x_n = y_m < x_n^{(2)} < \min\{x_{n+1}, y_{m+1}\},$$

where x_{n-1} and x_{n+1} (respectively, y_{m-1}, y_{m+1}) denote the preceding and subsequent points of x_n in S_X (respectively, of y_m in S_Y), existing because since both S_X and S_Y have no accumulation points. Let us use the following notation:

$$\begin{aligned} S_X^a &= \{x_n \in S_X : P(X = x_n, Y = x_n) = 0\}. \\ S_X^b &= \{x_n \in S_X : P(X = x_n, Y = x_n) > 0\}. \end{aligned}$$

Then, $S_X = S_X^a \cup S_X^b$. We define the random variable \tilde{X} whose support is given by:

$$S_{\tilde{X}} = \{x_n \in S_X^a\} \cup \{x_n^{(1)}, x_n^{(2)} : x_n \in S_X^b\}.$$

The joint probability of \tilde{X} and Y is given by:

$$\begin{aligned} P(\tilde{X} = x_n, Y = y_m) &= p_{n,m} \quad \text{if } x_n \in S_X^a. \\ P(\tilde{X} = x_n^{(1)}, Y = y_m) &= P(\tilde{X} = x_n^{(2)}, Y = y_m) = \frac{1}{2}p_{n,m} \quad \text{if } x_n \in S_X^b. \end{aligned}$$

By definition, $P(\tilde{X} = Y) = 0$. Then:

$$\begin{aligned}
Q(\tilde{X}, Y) &= P(\tilde{X} > Y) = \sum_{x \in S_{\tilde{X}}} P(\tilde{X} > Y \mid \tilde{X} = x) \\
&= \sum_{x_n \in S_X^a} P(\tilde{X} > Y \mid \tilde{X} = x_n) + \sum_{x_n \in S_X^b} P(\tilde{X} > Y \mid \tilde{X} = x_n^{(1)}) \\
&\quad + P(\tilde{X} > Y \mid \tilde{X} = x_n^{(2)}) \\
&= \sum_{x_n \in S_X^a} P(X > Y \mid X = x_n) + \sum_{x_n \in S_X^b} \frac{1}{2} P(X > Y \mid X = x_n) \\
&\quad + \frac{1}{2} \sum_{x_n \in S_X^b} \left(P(X > Y \mid X = x_n) + P(X = x_n, Y = x_n) \right) \\
&= \sum_{x_n \in S_X^a} P(X > Y \mid X = x_n) + \sum_{x_n \in S_X^b} \left(P(X > Y \mid X = x_n) \right. \\
&\quad \left. + \frac{1}{2} P(X = x_n, Y = x_n) \right) \\
&= \sum_{x_n \in S_X} P(X > Y \mid X = x_n) + \frac{1}{2} \sum_{x_n \in S_X^a} P(X = x_n, Y = x_n) \\
&= P(X > Y) + \frac{1}{2} P(X = Y) = Q(X, Y). \quad \blacksquare
\end{aligned}$$

This lemma allows us to establish the following theorem.

Theorem 3.42 *Let X and Y be two real-valued discrete random variables on the same probability space, whose supports have no accumulation points and such that $P(X = Y) > 0$. Then $X \succ_{\text{SP}} Y$ if and only if it is possible to find a random variable \tilde{X} in the conditions of Lemma 3.41 such that $\inf \text{Me}(\tilde{X} - Y) > 0$.*

Proof: Applying the previous lemma it is possible to build another random variable \tilde{X} such that $Q(\tilde{X}, Y) = Q(X, Y)$, $P(\tilde{X} = Y) = 0$, and if $P(X = Y \mid X = x) = 0$, then $P(\tilde{X} = x) = P(X = x)$.

Therefore, as $P(\tilde{X} = Y) = 0$, by Theorem 3.40 it holds that $\tilde{X} \succ_{\text{SP}} Y$ if and only if $\inf \text{Me}(\tilde{X} - Y) \geq 0$. But since $Q(\tilde{X}, Y) = Q(X, Y)$, it holds that $X \succ_{\text{SP}} Y$ if and only if $\inf \text{Me}(\tilde{X} - Y) \geq 0$. \blacksquare

3.2 Relationship between stochastic dominance and statistical preference

In this section we shall study the relationships between first degree stochastic dominance and statistical preference for real-valued random variables.

We recall once more that stochastic dominance only uses the marginal distributions of the variables compared. As we have seen in Subsection 2.1.2, every joint cumulative distribution function is the copula of the marginal cumulative distribution functions. For this reason, as we have already done in the previous subsection, we focus on different situations: independent, comonotonic and countermonotonic random variables, and random variables coupled by an Archimedean copula.

Before starting with the main results, we are going to show that in general, first degree stochastic dominance does not imply statistical preference.

Example 3.43 Consider the random variables X and Y whose joint mass probability function is given by:

$X \backslash Y$	0	1	2
0	0.2	0.15	0
1	0	0.2	0.15
2	0.2	0	0.1

Then, the marginal cumulative distribution functions of X and Y are defined by:

	$t < 0$	$t \in [0, 1)$	$t \in [1, 2)$	$t \geq 2$
$F_X(t)$	0	0.35	0.7	1
$F_Y(t)$	0	0.4	0.75	1

It follows that $X \succeq_{\text{FSD}} Y$ since $F_X \leq F_Y$. However, $X \not\succeq_{\text{SP}} Y$ since:

$$\begin{aligned} Q(X, Y) &= P(X > Y) + \frac{1}{2}P(X = Y) \\ &= P(X = 2, Y = 0) + \frac{1}{2} \left(P(X = 0, Y = 0) + P(X = 1, Y = 1) \right. \\ &\quad \left. + P(X = 2, Y = 2) \right) = 0.2 + \frac{1}{2}(0.2 + 0.2 + 0.1) = 0.45. \end{aligned}$$

Thus, $X \succeq_{\text{FSD}} Y$ does not imply $X \succeq_{\text{SP}} Y$. ♦

Furthermore, since $X \succeq_{\text{FSD}} Y$ implies $X \succeq_{n\text{SD}} Y$ for any $n \geq 2$, the previous example also shows that $X \succeq_{n\text{SD}} Y$ does not imply $X \succeq_{\text{SP}} Y$ for any $n \geq 2$.

In the following subsections, we will find sufficient conditions for the implication $X \succeq_{\text{FSD}} \Rightarrow X \succeq_{\text{SP}} Y$.

3.2.1 Independent random variables

We start by proving that first degree stochastic dominance implies statistical preference for independent random variables. For this aim, take into account that, when $X \succeq_{\text{FSD}} Y$, Theorem (2.10) assures that $E[u(X)] \geq E[u(Y)]$ for any increasing function u . In particular, if we consider $u = F_Y$, which is an increasing function, it holds that $E[F_Y(X)] \geq E[F_Y(Y)]$. This will be an interesting fact in order to prove the next result.

Theorem 3.44 *Let X and Y be two real-valued independent random variables. Then $X \succeq_{\text{FSD}} Y$ implies $X \succeq_{\text{SP}} Y$.*

Proof: Using Lemma 2.20, it suffices to prove that

$$P(X \geq Y) \geq P(Y \geq X).$$

Since X and Y are independent, by Lemma 3.11 it is equivalent to prove that:

$$E[F_Y(X)] \geq E[F_X(Y)].$$

Moreover, since $X \succeq_{\text{FSD}} Y$, $F_X \leq F_Y$, and therefore $E[F_X(Y)] \leq E[F_Y(Y)]$. Thus, it suffices to prove that

$$E[F_Y(X)] \geq E[F_Y(Y)],$$

and this inequality holds because $X \succeq_{\text{FSD}} Y$ and then $E[u(X)] \geq E[u(Y)]$ for every increasing function u . ■

With a similar proof it is possible to establish that the implication holds even when one of the variables strictly dominates the other one. Let us introduce a preliminary lemma.

Lemma 3.45 *Let X and Y be two independent real-valued random variables such that $X \succ_{\text{FSD}} Y$. Then, if $P(Y = t) = 0$ for any t such that $F_X(t) < F_Y(t)$, there exists an interval $[a, b]$ such that $P(Y \in [a, b]) > 0$ and $F_X(t) < F_Y(t)$ for any $t \in [a, b]$.*

Proof: Let t_0 be a point such that $F_X(t_0) < F_Y(t_0)$. Since both F_X and F_Y are right-continuous,

$$\lim_{\varepsilon \rightarrow 0} F_Y(t_0 + \varepsilon) = F_Y(t_0) > F_X(t_0) = \lim_{\varepsilon \rightarrow 0} F_X(t_0 + \varepsilon).$$

Then, there is $\varepsilon > 0$ such that:

$$F_X(t_0 + \varepsilon) \leq F_X(t_0) + \frac{F_Y(t_0) - F_X(t_0)}{2} < F_Y(t_0).$$

Considering $\delta = \frac{F_Y(t_0) - F_X(t_0)}{2} > 0$, then $F_Y(t) - F_X(t) \geq \delta > 0$ for any $t \in [t_0, t_0 + \varepsilon]$. We have thus proven that there exists an interval $[a, b]$ such that $F_Y(t) - F_X(t) \geq \delta > 0$ for $t \in [a, b]$. Now, without loss of generality, we can assume that $F_Y(a - \varepsilon) < F_Y(a)$ for any $\varepsilon > 0$ (otherwise, since F_Y is right-continuous, take the point $a^* = \inf\{t : F_Y(t) = F_Y(a)\}$). Then, since $P(Y = a) = 0$, there exists $\varepsilon > 0$ such that $F_Y(t) - F_X(t) \geq \delta > 0$ for any $t \in [a - \varepsilon, b]$. Furthermore:

$$P(Y \in [a - \varepsilon, b]) \geq P(Y \in [a - \varepsilon, a]) \geq P(Y \in (a - \varepsilon, a]) = F_Y(a) - F_Y(a - \varepsilon) > 0,$$

and this completes the proof. ■

The following result had already been established in [14, Proposition 15.3.5]. However, the authors only gave a proof for continuous random variables. Here, we provide a proof for any pair of random variables X and Y .

Proposition 3.46 *Let X and Y be two real-valued independent random variables. Then, $X \succ_{\text{FSD}} Y$ implies $X \succ_{\text{SP}} Y$.*

Proof: We have proven in Theorem 3.44 that $E[F_Y(X)] \geq E[F_Y(Y)]$ when $X \succeq_{\text{FSD}} Y$. Then, if we prove that $E[F_X(Y)] < E[F_Y(Y)]$ we would obtain that:

$$P(X \geq Y) = E[F_Y(X)] \geq E[F_Y(Y)] > E[F_X(Y)] = P(Y \geq X),$$

and consequently $X \succ_{\text{SP}} Y$.

Let us prove that if $X \succ_{\text{FSD}} Y$, then $E[F_X(Y)] < E[F_Y(Y)]$. By hypothesis, $F_X(t) \leq F_Y(t)$ for every t , and there is t_0 such that $F_X(t_0) < F_Y(t_0)$.

Let us consider two cases. On the one hand, let us assume that $P(Y = t_0) > 0$. In such a case:

$$\begin{aligned} E[F_X(Y)] &= \int F_X dF_Y = \int_{\mathbb{R} \setminus \{t_0\}} F_X dF_Y + \int_{\{t_0\}} F_X dF_Y \\ &\leq \int_{\mathbb{R} \setminus \{t_0\}} F_Y dF_Y + P(Y = t_0) F_X(t_0) \end{aligned}$$

On the other hand, assume that there is not t_0 satisfying both $F_X(t_0) < F_Y(t_0)$ and $P(Y = t_0) > 0$. Applying the previous lemma, there is an interval $[a, b]$ such that $F_Y(t) - F_X(t) \geq \delta > 0$ and $P(Y \in [a, b]) > 0$. Then:

$$\begin{aligned} E[F_X(Y)] &= \int F_X dF_Y = \int_{\mathbb{R} \setminus [a, a+\varepsilon]} F_X dF_Y + \int_{[a, a+\varepsilon]} F_X dF_Y \\ &\leq \int_{\mathbb{R} \setminus [a, a+\varepsilon]} F_Y dF_Y + \int_{[a, a+\varepsilon]} (F_Y - \delta) dF_Y \\ &= \int F_Y dF_Y - \delta P(Y \in [a, a + \varepsilon]) < E[F_Y(Y)]. \quad \blacksquare \end{aligned}$$

A similar result was proven in [210] for probability dominance (see Remark 2.22); nevertheless, that result was only valid for continuous random variables.

3.2.2 Continuous comonotonic and countermonotonic random variables

Let X and Y be two random variables with respective cumulative distribution functions F_X and F_Y , and respective density functions f_X and f_Y .

First of all, let us study the relationship between first degree stochastic dominance and statistical preference for comonotonic random variables.

Theorem 3.47 *Let X and Y be two real-valued comonotonic and continuous random variables. If $X \succeq_{\text{FSD}} Y$, then $X \succeq_{\text{SP}} Y$.*

Proof: In Corollary 3.17 we have seen that $X \succeq_{\text{SP}} Y$ if and only if

$$\int_{x:F_X(x) < F_Y(x)} (f_X(x) + f_Y(x))dx + \frac{1}{2} \int_{x:F_X(x)=F_Y(x)} (f_X(x) + f_Y(x))dx \geq 1.$$

However, by hypothesis $F_X(x) \leq F_Y(x)$ for any $x \in \mathbb{R}$. Then, $\{x : F_X(x) \leq F_Y(x)\} = \mathbb{R}$, and therefore:

$$\begin{aligned} & \int_{x:F_X(x) < F_Y(x)} (f_X(x) + f_Y(x))dx + \frac{1}{2} \int_{x:F_X(x)=F_Y(x)} (f_X(x) + f_Y(x))dx \\ &= \int_{x:F_X(x) \leq F_Y(x)} (f_X(x) + f_Y(x))dx - \frac{1}{2} \int_{x:F_X(x)=F_Y(x)} (f_X(x) + f_Y(x))dx \\ &= \int_{\mathbb{R}} (f_X(x) + f_Y(x))dx - \frac{1}{2} \int_{x:F_X(x)=F_Y(x)} (f_X(x) + f_Y(x))dx \\ &\geq \int_{\mathbb{R}} (f_X(x) + f_Y(x))dx - 1 = 2 - 1 = 1. \end{aligned}$$

Thus, X is statistically preferred to Y . ■

Proposition 3.46 assures that for independent random variables, when first degree stochastic dominance holds in the strict sense, statistical preference is also strict. As we shall see, this also holds for continuous and comonotonic real-valued random variables. In order to establish this, we give first the following lemma.

Lemma 3.48 *Let X and Y be two continuous real-valued random variables. Then, if $X \succ_{\text{FSD}} Y$, there exists an interval $[a, b]$ such that $F_X(t) < F_Y(t)$ for any $t \in [a, b]$ and $P(X \in [a, b]) > 0$.*

Proof: From the proof of Lemma 3.45 we deduce that there is an interval $[a, b]$ such that $F_Y(t) - F_X(t) \geq \delta > 0$ for any $t \in [a, b]$. Since F_X is continuous, there is $\varepsilon > 0$ such that $F_X(a - \varepsilon) < F_X(a)$ and $F_Y(t) - F_X(t) \geq \frac{\delta}{2} > 0$ for any $t \in [a - \varepsilon, b]$. Then:

$$P(X \in [a - \varepsilon, b]) \geq P(X \in [a - \varepsilon, a]) \geq F_X(a) - F_X(a - \varepsilon) > 0. \quad \blacksquare$$

Proposition 3.49 *Let X and Y be two real-valued comonotonic and continuous random variables. If $X \succ_{\text{FSD}} Y$, then $X \succ_{\text{SP}} Y$.*

Proof: On the one hand, since $X \succ_{\text{FSD}} Y$, then $X \succeq_{\text{FSD}} Y$, and consequently $X \succeq_{\text{SP}} Y$. According to the previous lemma, there is an interval $[a, b]$ such that $F_Y(t) - F_X(t) \geq \delta > 0$ for any $t \in [a, b]$ and $P(X \in [a, b]) > 0$. By Lemma 2.20, $X \succ_{\text{SP}} Y$ is equivalent to $P(X > Y) > P(Y > X)$, and from Proposition 3.16 this is equivalent to:

$$\int_{x:F_X(x) < F_Y(x)} f_X(x) dx > \int_{x:F_Y(x) < F_X(x)} f_Y(x) dx.$$

Now, take into account that the second part of the previous equation equals 0, since $\{x : F_Y(x) < F_X(x)\} = \emptyset$. In addition:

$$\int_{x:F_X(x) < F_Y(x)} f_X(x) dx \geq \int_{[a,b]} f_X(x) dx = P(X \in [a, b]) > 0.$$

Thus, we conclude that $X \succ_{\text{SP}} Y$. \blacksquare

When the random variables are countermonotonic, the relationship between the (non-strict) first degree stochastic dominance and the (non-strict) statistical preference also holds.

Theorem 3.50 *Let X and Y be two real-valued countermonotonic and continuous random variables. If $X \succeq_{\text{FSD}} Y$, then $X \succeq_{\text{SP}} Y$.*

Proof: In Proposition 3.19 we have seen that $X \succeq_{\text{SP}} Y$ if and only if $F_Y(u) \geq F_X(u)$, where u is one point such that $F_Y(u) + F_X(u) = 1$. However, since $X \succeq_{\text{FSD}} Y$, it holds that $F_X(x) \leq F_Y(x)$ for every $x \in \mathbb{R}$. In particular, it also holds that $F_X(u) \leq F_Y(u)$. \blacksquare

Although it seems intuitive that the same relationship holds with respect to the strict preferences, this is not the case for countermonotonic continuous random variables. To see this, it suffices to consider the countermonotonic random variables X and Y whose cumulative distribution functions of X and Y are defined by:

$$F_X(t) = \begin{cases} 0 & \text{if } t < 0. \\ t & \text{if } t \in [0, 1]. \\ 1 & \text{if } t > 1. \end{cases} \quad (3.15)$$

$$F_Y(t) = \begin{cases} 0 & \text{if } t < -0.1. \\ \frac{1}{2}t + 0.05 & \text{if } t \in [-0.1, 0.1]. \\ t & \text{if } t \in [0.1, 1]. \\ 1 & \text{if } t > 1. \end{cases} \quad (3.16)$$

Since $F_X(t) = F_Y(t)$ for any $t \notin (-0.1, 0.1)$ and $F_X(t) < F_Y(t)$ for $t \in (-0.1, 0.1)$, it holds that $X \succ_{\text{FSD}} Y$, but $X \equiv_{\text{SP}} Y$, since $F_X(u) + F_Y(u) = 1$ for $u = \frac{1}{2}$ and:

$$Q(X, Y) = F_Y(u) = F_Y(0.5) = \frac{1}{2}.$$

$$Q(Y, X) = F_X(u) = F_X(0.5) = \frac{1}{2}.$$

3.2.3 Discrete comonotonic and countermonotonic random variables with finite supports

Let us now assume that X and Y are discrete real-valued random variables with finite support. Then, when these random variables are comonotonic, we obtain the following result:

Theorem 3.51 *If X and Y are two real-valued comonotonic and discrete random variables with finite supports, then $X \succeq_{\text{FSD}} Y \Rightarrow X \succeq_{\text{SP}} Y$.*

Proof: Using Remark 3.24, we can assume w.l.o.g. that X and Y are defined in $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \dots, \omega_n\}$, by $X(\omega_i) = x_i$ and $Y(\omega_i) = y_i$, where $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for any $i = 1, \dots, n-1$, and also:

$$P(X = x_i, Y = y_i) = P(X = x_i) = P(Y = y_i) \text{ for any } i = 1, \dots, n.$$

Moreover, using Proposition 3.25, $X \succeq_{\text{SP}} Y$ if and only if

$$\sum_{i: x_i > y_i} P(X = x_i) \geq \sum_{i: x_i < y_i} P(X = x_i).$$

Let us show that $\{i : x_i < y_i\} = \emptyset$ when $X \succeq_{\text{FSD}} Y$. Assume that there exists k such that $X(\omega_k) = x_k < y_k = Y(\omega_k)$. Then:

$$\begin{aligned} F_X(x_k) &= P(X \leq X(x_k)) \geq P(\{\omega_1, \dots, \omega_k\}). \\ F_Y(x_k) &= P(Y \leq X(x_k)) \leq P(\{\omega_1, \dots, \omega_{k-1}\}), \end{aligned}$$

where last inequality holds since $\omega_k \notin \{Y \leq X(x_k)\}$ because $Y(\omega_k) > X(\omega_k)$. Now, since $X \succeq_{\text{FSD}} Y$, it holds that $F_X(x_k) \leq F_Y(x_k)$:

$$P(\{\omega_1, \dots, \omega_k\}) \leq F_X(x_k) \leq F_Y(x_k) \leq P(\{\omega_1, \dots, \omega_{k-1}\}).$$

This implies that $P(\{\omega_k\}) = P(\{X = x_k\}) = 0$, but a contradiction arises since $P(\{\omega_k\}) > 0$. Then, we conclude that $\{i : x_i > y_i\} = \emptyset$, and consequently:

$$\sum_{i: x_i > y_i} P(X = x_i) \geq 0 = \sum_{i: x_i < y_i} P(X = x_i).$$

Thus, $X \succeq_{\text{SP}} Y$. ■

Now, it only remains to see that, as for continuous random variables, strict stochastic dominance implies strict statistical preference.

Proposition 3.52 *Let X and Y be two real-valued discrete and countermonotonic random variables with finite supports. Then, $X \succ_{\text{FSD}} Y$ implies $X \succ_{\text{SP}} Y$.*

Proof: It is obvious that $X \succ_{\text{FSD}} Y$ implies $X \succeq_{\text{FSD}} Y$, and then, applying the previous theorem, $X \succeq_{\text{SP}} Y$ because $\{i : x_i < y_i\} = \emptyset$. Then, in order to prove that $X \succ_{\text{SP}} Y$ it is enough to see that $\{i : x_i > y_i\} \neq \emptyset$, that is, there is some k such that $x_k > y_k$.

Since $X \succ_{\text{FSD}} Y$, there is some k such that $F_X(y_k) < F_Y(y_k)$. Assume ex-absurdo that $\{i : x_i > y_i\} = \emptyset$, so $x_i = y_i$ for any $i = 1, \dots, n$. Since $x_i = y_i$ and $P(X = x_i) = P(Y = y_i) = P(Y = x_i)$, X and Y are equally distributed, and then $X \equiv_{\text{FSD}} Y$, a contradiction. ■

Finally, let us consider discrete countermonotonic random variables with finite supports, and let us see that, in that case, first degree stochastic dominance also implies statistical preference.

Theorem 3.53 *Let X and Y be two real-valued discrete and countermonotonic random variables with finite supports. Then, $X \succeq_{\text{FSD}} Y$ implies $X \succeq_{\text{SP}} Y$.*

Proof: From remark 3.32, without loss of generality we can assume that X and Y are defined on $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \dots, \omega_n\}$, by $X(\omega_i) = x_i$ and $Y(\omega_i) = y_{n-i+1}$, where $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for any $i = 1, \dots, n-1$, and also:

$$P(X = x_i, Y = y_i) = P(X = x_i) = P(Y = y_{n-i+1}) \text{ for any } i = 1, \dots, n.$$

Furthermore, we can also assume that

$$\max(|X(\omega_i) - X(\omega_{i+1})|, |Y(\omega_i) - Y(\omega_{i+1})|) > 0 \text{ for any } i = 1, \dots, n-1;$$

and that there exists, at most, one element k such that $X(\omega_k) = Y(\omega_k)$.

In order to prove that $X \succeq_{\text{FSD}} Y \Rightarrow X \succeq_{\text{SP}} Y$ we consider two cases:

- Assume $X(\omega_i) \neq Y(\omega_i)$ for any $i = 1, \dots, n$ and denote $k = \max\{i : X(\omega_i) < Y(\omega_i)\}$. Then, by Proposition 3.33, $X \succeq_{\text{SP}} Y$ if and only if:

$$P(\{\omega_1\}) + \dots + P(\{\omega_k\}) \leq P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\}).$$

Since $X \succeq_{\text{FSD}} Y$, $F_X \leq F_Y$. Then, taking $\varepsilon = \frac{Y(\omega_k) - X(\omega_k)}{2} > 0$, it holds that:

$$\begin{aligned} F_X(X(\omega_k)) &= P(X \leq X(\omega_k)) \geq P(\{\omega_1, \dots, \omega_k\}). \\ F_Y(X(\omega_k)) &\leq F_Y(Y(\omega_k) - \varepsilon) = P(Y \leq Y(\omega_k) - \varepsilon) \leq P(\{\omega_{k+1}, \dots, \omega_n\}). \end{aligned}$$

- Assume that there is (an unique) k such that $X(\omega_k) = Y(\omega_k)$. Then:

$$\begin{aligned} F_X(X(\omega_{k-1})) &= P(X' \leq X'(\omega_{k-1})). \\ F_Y(X(\omega_{k-1})) &= P(Y' \leq Y'(\omega_{k-1})). \end{aligned}$$

Since $X(\omega_{k-1}) < Y(\omega_{k-1})$, $\omega_{k-1} \notin \{Y \leq X(\omega_{k-1})\}$, and this implies that $\{Y \leq X(\omega_{k-1})\} \subseteq \{\omega_k, \omega_{k+1}, \dots, \omega_n\}$. Furthermore, $\{X \leq X(\omega_{k-1})\} \supseteq \{\omega_1, \dots, \omega_{k-1}\}$, and then

$$F_X(X(\omega_{k-1})) \geq P(\{\omega_1\}) + \dots + P(\{\omega_{k-1}\}).$$

We consider two cases:

- Assume that $Y(\omega_k) = X(\omega_{k-1})$. Then $X(\omega_k) = Y(\omega_k) = X(\omega_{k-1})$, and this implies that $\omega_k \in \{X \leq X(\omega_{k-1})\}$. Then:

$$\begin{aligned} F_X(X(\omega_{k-1})) &\geq P(\{\omega_1\}) + \dots + P(\{\omega_{k-1}\}) + P(\{\omega_k\}). \\ F_Y(Y(\omega_{k-1})) &= P(\{\omega_k\}) + P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\}). \end{aligned}$$

Using that $X \succeq_{\text{FSD}} Y$,

$$P(\{\omega_1\}) + \dots + P(\{\omega_{k-1}\}) \geq P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\}).$$

Applying Proposition 3.33, $X \succeq_{\text{SP}} Y$.

- On the other hand, if $Y(\omega_k) \not\leq X(\omega_{k-1})$, then it holds that $\{Y \leq X(\omega_{k-1})\} \subseteq \{\omega_{k+1}, \dots, \omega_n\}$. Hence:

$$F_Y(X(\omega_{k-1})) = P(Y \leq X(\omega_{k-1})) \leq P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\})$$

and, since $F_X \leq F_Y$ because $X \succeq_{\text{FSD}} Y$, it holds that:

$$\begin{aligned} P(\{\omega_{k+1}\}) + \dots + P(\{\omega_n\}) &\geq P(Y \leq X(\omega_{k-1})) = F_Y(X(\omega_{k-1})) \\ &\geq F_X(X(\omega_{k-1})) = P(Y \leq X(\omega_1)) \\ &\geq P(\{\omega_{k+1}\}) + \dots + P(\{\omega_{k-1}\}). \end{aligned}$$

By Proposition 3.33, $X \succeq_{\text{SP}} Y$. ■

Unsurprisingly, strict first degree stochastic dominance does not imply strict statistical preference, as we can see in the following example:

Example 3.54 Consider the countermonotonic random variables X and Y defined by:

X, Y	0	1	2
P_X	0	0.2	0.8
P_Y	0.1	0.1	0.8

For these variables, $X \succ_{\text{FSD}} Y$. From Remark 3.32 we can assume that X and Y are defined in the probability space $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \dots, \omega_5\}$, and such that:

$P(\{\omega_i\})$	0.2	0.6	0.1	0.1
Ω	ω_1	ω_2	ω_3	ω_4
X	1	2	2	2
Y	2	2	1	0

Then, $Q(X, Y) = 0.5$, and we conclude that $X \equiv_{\text{SP}} Y$. ◆

3.2.4 Random variables coupled by an Archimedean copula

In this subsection we consider two continuous random variables X and Y , with respective cumulative distribution functions F_X , F_Y and with respective density functions f_X and f_Y . We assume that the random variables are coupled by an Archimedean copula C , generated by the twice differentiable function φ .

First of all, we consider the case of a strict Archimedean copula. In that case, we also obtain that first degree stochastic dominance implies that statistical preference.

Theorem 3.55 *Let X and Y be two real-valued continuous random variables coupled by a strict Archimedean copula C generated by the twice differentiable function φ . Then, $X \succeq_{\text{FSD}} Y$ implies $X \succeq_{\text{SP}} Y$.*

Proof: From Theorem 3.34, $X \succeq_{\text{SP}} Y$ if and only if:

$$E \left[\left((\varphi^{-1})' (\varphi(F_X(X)) + \varphi(F_Y(X))) - (\varphi^{-1})' (2\varphi(F_X(X))) \right) \varphi'(F_X(X)) \right] \geq 0,$$

or equivalently, if

$$\begin{aligned} \int_{-\infty}^{\infty} (\varphi^{-1})' (\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx \\ \geq \int_{-\infty}^{\infty} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx. \end{aligned}$$

This inequality holds because

$$\begin{aligned} X \succeq_{\text{FSD}} Y &\Rightarrow F_X(x) \leq F_Y(x) \\ &\Rightarrow \varphi(F_X(x)) \geq \varphi(F_Y(x)) && (\varphi \text{ is decreasing}) \\ &\Rightarrow 2\varphi(F_X(x)) \geq \varphi(F_X(x)) + \varphi(F_X(x)) \\ &\Rightarrow \begin{cases} (\varphi^{-1})' (2\varphi(F_X(x))) \geq \\ (\varphi^{-1})' (\varphi(F_X(x)) + \varphi(F_X(x))) \end{cases} && ((\varphi^{-1})' \text{ is increasing}) \\ &\Rightarrow \begin{cases} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X \leq \\ (\varphi^{-1})' (\varphi(F_X(x)) + \varphi(F_X(x))) \varphi'(F_X(x)) f_X \end{cases} && (\varphi' \leq 0.) \end{aligned}$$

Therefore, X is statistically preferred to Y . ■

Remark 3.56 *When applying the previous result to the product copula, we obtain that for continuous and independent random variables, $X \succeq_{\text{FSD}} Y \Rightarrow X \succeq_{\text{SP}} Y$. This is not new for us, since Theorem 3.44 states that this relation holds, not only for continuous, but any kind of independent random variables.♦*

Let us now investigate if such relationship also holds for the strict preference. For this aim, we consider this preliminary lemma.

Lemma 3.57 *Let X and Y be two continuous random variables such that $X \succ_{\text{FSD}} Y$. Then, there exists an interval $[a, b]$ such that $F_X(t) < F_Y(t)$ for any $t \in [a, b]$, and also $P(X \in [a, b]) > 0$ and*

$$(\varphi^{-1})'(\varphi(F_X(t)) + \varphi(F_Y(t)))\varphi'(F_X(t)) - (\varphi^{-1})'(2\varphi(F_X(t)))\varphi'(F_X(t)) \geq \delta > 0$$

for any $t \in [a, b]$.

Proof: We have proven in Lemma 3.48 that there exists an interval $[a, b]$ such that $F_Y(t) - F_X(t) \geq \delta > 0$ for any $t \in [a, b]$ and $P(X \in [a, b]) > 0$. Then, there is a subinterval $[a_1, b_1]$ of $[a, b]$ where F_X is strictly increasing.

Now, following the same steps than in Theorem 3.55 we obtain that:

$$\begin{aligned} F_X(t) < F_Y(t) \text{ for any } t \in [a, b] &\Rightarrow \\ (\varphi^{-1})'(\varphi(F_X(t)) + \varphi(F_Y(t)))\varphi'(F_X(t)) &> \\ (\varphi^{-1})'(2\varphi(F_X(t)))\varphi'(F_X(t)) &\text{ for any } t \in [a_1, b_1]. \end{aligned}$$

Consider $t \in [a_1, b_1]$ and let

$$\varepsilon = (\varphi^{-1})'(\varphi(F_X(t)) + \varphi(F_Y(t)))\varphi'(F_X(t)) - (\varphi^{-1})'(2\varphi(F_X(t)))\varphi'(F_X(t)) > 0.$$

Then, there is a subinterval $[a_2, b_2]$ of $[a_1, b_1]$ such that

$$(\varphi^{-1})'(\varphi(F_X(t)) + \varphi(F_Y(t)))\varphi'(F_X(t)) - (\varphi^{-1})'(2\varphi(F_X(t)))\varphi'(F_X(t)) \geq \frac{\varepsilon}{2} > 0.$$

Furthermore, since F_X is strictly increasing in $[a, b]$, it is also strictly increasing in $[a_2, b_2]$, and then $P(X \in [a_2, b_2]) > 0$. \blacksquare

Proposition 3.58 *Consider two real-valued continuous random variables X and Y coupled by a strict Archimedean copula C generated by φ . Then, $X \succ_{\text{FSD}} Y$ implies $X \succ_{\text{SP}} Y$.*

Proof: We have to prove that:

$$\begin{aligned} \int_{-\infty}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x)))\varphi'(F_X(x))f_X(x)dx \\ > \int_{-\infty}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x)))\varphi'(F_X(x))f_X(x)dx. \end{aligned}$$

Since X and Y are continuous, if $X \succ_{\text{FSD}} Y$, then $X \succeq_{\text{FSD}} Y$, and consequently $X \succeq_{\text{SP}} Y$ by Theorem 3.55. Taking into account the previous lemma, there exists an interval $[a, b]$ such that $P(X \in [a, b]) > 0$ and:

$$(\varphi^{-1})'(\varphi(F_X(t)) + \varphi(F_Y(t)))\varphi'(F_X(t)) - (\varphi^{-1})'(2\varphi(F_X(t)))\varphi'(F_X(t)) \geq \delta > 0$$

for any $t \in [a, b]$. Then:

$$\begin{aligned}
& \int_{-\infty}^{\infty} (\varphi^{-1})' (\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx \\
& \geq \int_{\mathbb{R}-[a,b]} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx \\
& \quad + \int_{[a,b]} (\varphi^{-1})' (\varphi(F_X(x)) + \varphi(F_Y(x))) \varphi'(F_X(x)) f_X(x) dx \\
& > \int_{[a,b]} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx \\
& \quad + \int_{[a,b]} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx + \int_{[a,b]} \frac{\varepsilon}{2} f_X(x) dx \\
& = \int_{-\infty}^{\infty} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx + \frac{\varepsilon}{2} P(X \in [a, b]) \\
& > \int_{-\infty}^{\infty} (\varphi^{-1})' (2\varphi(F_X(x))) \varphi'(F_X(x)) f_X(x) dx.
\end{aligned}$$

Consequently, $X \succ_{\text{SP}} Y$. ■

Remark 3.59 *As we have already mentioned, in the particular case where the strict Archimedean copula is the product, the relation $X \succ_{\text{FSD}} Y \Rightarrow X \succ_{\text{SP}} Y$ was already studied in Proposition 3.46. Such result states the relation not only for continuous, but for every kind of independent random variables.◆*

It only remains to study the case of nilpotent copulas. In order to do this, we are going to see the following lemma that assures that, over the assumption of $X \succeq_{\text{FSD}} Y$, the points \bar{x} and \tilde{x} , defined on Equations (3.12) and (3.13), respectively, satisfy $\bar{x} \leq \tilde{x}$.

Lemma 3.60 *Let X and Y be two real-valued continuous random variables coupled by a nilpotent Archimedean copula C generated by φ . If $X \succeq_{\text{FSD}} Y$, then it holds that $\bar{x} \leq \tilde{x}$.*

Proof: First of all, recall that:

$$\begin{aligned}
\tilde{x} &= \inf\{x \mid 2\varphi(F_X(x)) \leq \varphi(0)\}, \\
\bar{x} &= \inf\{x : y_x < x\} \text{ and} \\
y_x &= \inf\{y \mid \varphi(F_X(x)) + \varphi(F_Y(y)) \in [0, \varphi(0)]\} \text{ for any } x \in \mathbb{R}.
\end{aligned}$$

Assume that $\tilde{x} < \bar{x}$. Then there exists a point t^* such that $\tilde{x} < t^* < \bar{x}$ and $y_{t^*} > t^*$. Moreover, from the hypothesis $X \succeq_{\text{FSD}} Y$, it holds that

$$F_X(t) \leq F_Y(t) \Rightarrow \varphi(F_X(t)) \geq \varphi(F_Y(t)) \quad \forall t \in \mathbb{R}.$$

As $\tilde{x} < t^*$, we know that $2\varphi(F_X(t^*)) < \varphi(0)$. Therefore, we have that:

$$\varphi(F_X(t^*)) + \varphi(F_Y(t^*)) \leq 2\varphi(F_X(t^*)) < \varphi(0).$$

Then,

$$y_{t^*} = \inf\{y \mid \varphi(F_X(t^*)) + \varphi(F_Y(y)) < \varphi(0)\} \leq t^*.$$

Therefore, $y_{t^*} > t^* \geq y_{t^*}$, a contradiction. We conclude that $\tilde{x} \geq \bar{x}$. \blacksquare

Using this lemma we can prove that first degree stochastic dominance also implies statistical preference for continuous random variables coupled by a nilpotent Archimedean copula.

Theorem 3.61 *If X and Y are two real-valued continuous random variables coupled by a nilpotent Archimedean copula whose generator φ is twice differentiable such that $\varphi'' \neq 0$, then $X \succeq_{\text{FSD}} Y \Rightarrow X \succeq_{\text{SP}} Y$.*

Proof: From Lemma 3.60, $\bar{x} \leq \tilde{x}$. Furthermore, $F_X(x) \leq F_Y(x)$ for every $x \in \mathbb{R}$. Then, for every $x \geq \tilde{x}$:

$$\begin{aligned} X \succeq_{\text{FSD}} Y &\Rightarrow F_X(x) \leq F_Y(x) \\ &\Rightarrow \varphi(F_X(x)) \geq \varphi(F_Y(x)) && (\varphi \text{ is decreasing}) \\ &\Rightarrow 2\varphi(F_X(x)) \geq \varphi(F_X(x)) + \varphi(F_X(x)) \\ &\Rightarrow \begin{cases} (\varphi^{-1})'(2\varphi(F_X(x))) \geq \\ (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_X(x))) \end{cases} && ((\varphi^{-1})' \text{ is increasing}) \\ &\Rightarrow \begin{cases} (\varphi^{-1})'(2\varphi(F_X(x)))\varphi'(F_X(x))f_X \leq \\ (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_X(x)))\varphi'(F_X(x))f_X \end{cases} && (\varphi' \leq 0.) \end{aligned}$$

Therefore:

$$\begin{aligned} &\int_{\tilde{x}}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x)))\varphi'(F_X(x))f_X(x)dx \\ &\geq \int_{\tilde{x}}^{\infty} (\varphi^{-1})'(\varphi(F_X(x)) + \varphi(F_Y(x)))\varphi'(F_X(x))f_X(x)dx \\ &\geq \int_{\tilde{x}}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x)))\varphi'(F_X(x))f_X(x)dx \\ &\geq \int_{\tilde{x}}^{\infty} (\varphi^{-1})'(2\varphi(F_X(x)))\varphi'(F_X(x))f_X(x)dx. \end{aligned}$$

Applying Theorem 3.35, we deduce that $X \succeq_{\text{SP}} Y$. \blacksquare

Remark 3.62 *Note that this result is not applicable to the Łukasiewicz copula, since its generator is $\varphi_W(t) = 1 - t$, and then $\varphi''(t) = 0$. However, we have already seen in Theorem 3.50 that first degree stochastic dominance implies statistical preference for continuous and countermonotonic random variables. \blacklozenge*

As in the countermonotonic case, the relationship between the strict preferences does not hold. To see this, consider two continuous random variables X and Y whose cumulative distribution functions are defined in Equations (3.15) and (3.16). If we consider the generator $\varphi(t) = 2(1 - \sqrt{t})$, such that $\varphi(0) = 2$, there is not (x, y) in the set:

$$\{(x, y) : \varphi(F_X(x)) + \varphi(F_Y(y)) \in [0, \varphi(0)]\} \Rightarrow F_X(t) = F_Y(t),$$

such that either $x \leq 0.1$ or $y \leq 0.1$. Thus, whenever $f_{X,Y} > 0$, $f_{X,Y}$ is symmetric. Then, if (t, t) satisfies $\varphi(F_X(t)) + \varphi(F_Y(t)) \in [0, \varphi(0)]$, then $F_X(t) = F_Y(t)$. Consequently:

$$P(X > Y) = \int_{-\infty}^{\infty} \int_{-\infty}^x f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^x f_{X,Y}(y, x) dy dx = P(Y > X).$$

and we conclude X and Y are statistically indifferent.

3.2.5 Other relationships between stochastic dominance and statistical preference

In the previous subsection we have seen several conditions under which $X \succeq_{\text{FSD}} Y$ implies $X \succeq_{\text{SP}} Y$. Now, we analyze if there are other relationships between first and n -th degree stochastic dominance and statistical preference.

We start by proving that statistical preference does not imply neither first nor n -th degree stochastic dominance for any $n \geq 2$.

Remark 3.63 *There exist random variables X and Y such that:*

1. $X \succeq_{\text{SP}} Y$ but $X \not\succeq_{\text{nSD}} Y$, for every $n \geq 1$.
2. $X \succeq_{\text{nSD}} Y$ but $X \not\succeq_{\text{SP}} Y$, for every $n \geq 2$.
3. $X \succeq_{\text{FSD}} Y$ but $X \not\succeq_{\text{SP}} Y$.
4. $X \succeq_{\text{FSD}} Y$, $X \succeq_{\text{nSD}} Y$ for any $n \geq 2$ but $X \not\succeq_{\text{FSD}} Y$.

In Example 3.43 we gave two random variables such that $Y \succ_{\text{SP}} X$ but $X \succ_{\text{FSD}} Y$. Then, $X \succ_{\text{nSD}} Y$ for any $n \geq 1$ and therefore $Y \not\succeq_{\text{nSD}} X$ for any $n \geq 1$. Thus, this is an example where the first and third items hold.

Consider next random variables X and Y such that X follows a uniform distribution in the interval $(10, 11)$ and Y has the following density function:

$$f_Y(x) = \begin{cases} \frac{1}{25} & \text{if } 0 < x < 10, \\ \frac{3}{5} & \text{if } 11 < x < 12, \\ 0 & \text{otherwise.} \end{cases}$$

For these random variables it holds that:

$$Q(X, Y) = P(X > Y) = P(Y < 10) = \frac{2}{5} < \frac{1}{2},$$

and therefore $Y \succ_{\text{SP}} X$. However, on the one hand, it is trivial that neither $Y \succeq_{\text{FSD}} X$ nor $X \succeq_{\text{FSD}} Y$. Moreover, $X \succeq_{n\text{SD}} Y$ for every $n \geq 2$:

$$G_X^2(t) = \begin{cases} 0 & \text{if } t < 10. \\ \frac{(t-10)^2}{2} & \text{if } t \in [10, 11). \\ t - 10.5 & \text{if } t \geq 11. \end{cases}$$

$$G_Y^2(t) = \begin{cases} 0 & \text{if } t < 0. \\ \frac{t^2}{50} & \text{if } t \in [0, 10). \\ \frac{2}{5}t - 2 & \text{if } t \in [10, 11). \\ \frac{1}{10}(343 - 62t + 3t^2) & \text{if } t \in [11, 12). \\ t - 8.9 & \text{if } t \geq 12. \end{cases}$$

The graphs of these functions can be seen in Figure 3.1.

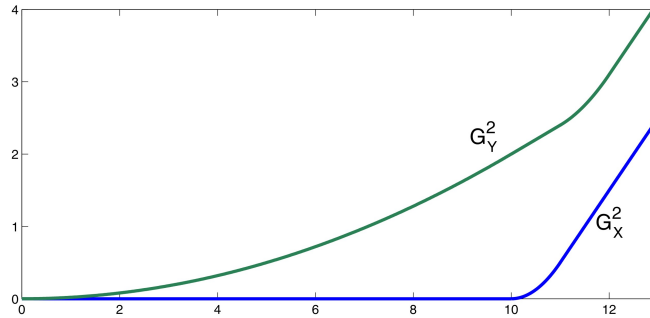


Figure 3.1: Graphics of the functions G_X^2 and G_Y^2 .

Then, $X \succeq_{\text{SSD}} Y$, and applying Equation (2.4), $X \succeq_{n\text{SD}} Y$ for every $n \geq 2$.

We have thus an example where $Y \succ_{\text{SP}} X$ and $X \succeq_{n\text{SD}} Y$ for every $n \geq 2$.

Let us see by means of an example that $X \succeq_{\text{SP}} Y$ and $X \succeq_{n\text{SD}} Y$ do not guarantee $X \succeq_{\text{FSD}} Y$. To see that, it is enough to consider the independent random variables X and Y defined by:

X	1	5	Y	0	10
P_X	$\frac{1}{2}$	$\frac{1}{2}$	P_Y	$\frac{9}{10}$	$\frac{1}{10}$

For these variables it holds that:

$$Q(X, Y) = P(X > Y) + \frac{1}{2}P(X = Y) = P(X > Y) = P(Y = 0) = \frac{9}{10} > \frac{1}{2}.$$

Thus $X \succeq_{\text{SP}} Y$. Furthermore, since the cumulative distribution functions are:

$$F_X(t) = \begin{cases} 0 & \text{if } t < 1, \\ \frac{1}{2} & \text{if } t \in [1, 5), \\ 1 & \text{if } t \geq 5. \end{cases} \quad F_Y(t) = \begin{cases} 0 & \text{if } t < 0, \\ \frac{9}{10} & \text{if } t \in [0, 10), \\ 1 & \text{if } t \geq 10. \end{cases}$$

the functions G_X^2 and G_Y^2 are:

$$G_X^2(t) = \begin{cases} 0 & \text{if } t < 1, \\ \frac{1}{2}(t-1) & \text{if } t \in [1, 5), \\ t-3 & \text{if } t \geq 5, \end{cases} \quad G_Y^2(t) = \begin{cases} 0 & \text{if } t < 0, \\ \frac{9}{10}t & \text{if } t \in [0, 10), \\ t-1 & \text{if } t \geq 10. \end{cases}$$

If we look at their graphical representations in Figure 3.2, we can see that $X \succeq_{\text{SSD}} Y$. However,

$$F_X(5) = 1 > \frac{9}{10} = F_Y(5),$$

whence X cannot stochastically dominate Y by first degree, i.e., $X \not\succeq_{\text{FSD}} Y$. ♦

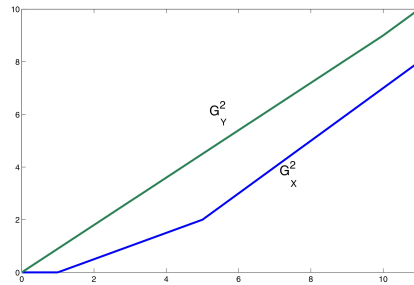


Figure 3.2: Graphics of the functions G_X^2 and G_Y^2 .

Our next Theorem summarises the main results of this paragraph.

Theorem 3.64 Let X and Y be two random variables. $X \succeq_{\text{FSD}} Y$ implies $X \succeq_{\text{SP}} Y$ under any of the following conditions:

- X and Y are independent.

- X and Y are continuous and comonotonic random variables.
- X and Y are continuous and countermonotonic random variables.
- X and Y are discrete and comonotonic random variables with finite supports.
- X and Y are discrete and countermonotonic random variables with finite supports.
- X and Y are continuous random variables coupled by an Archimedean copula.

The relationships between stochastic dominance and statistical preference under the conditions of the previous result are summarised in Figure 3.3.

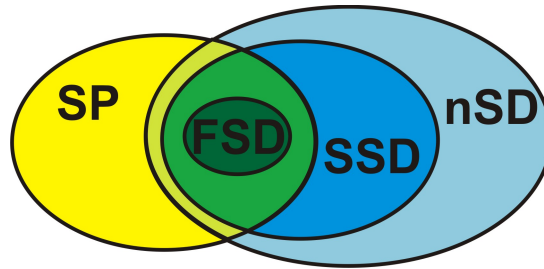


Figure 3.3: General relationship between stochastic dominance and statistical preference.

3.2.6 Examples on the usual distributions

In this subsection we shall study the conditions we must impose on the parameters of some of the most important parametric distributions in order to obtain statistical preference and stochastic dominance for independent random variables. We shall see that for some of them, stochastic dominance and statistical preference are equivalent. Some results in this sense have already been established in [56].

Discrete distributions under independence: Bernoulli

In the case of discrete distributions, we shall consider the Bernoulli distribution with parameter $p \in (0, 1)$, denoted by $\mathcal{B}(p)$, that takes the value 1 with probability p and the value 0 with probability $1 - p$.

Proposition 3.65 *Let X and Y be two independent random variables with distributions $X \equiv \mathcal{B}(p_1)$ and $Y \equiv \mathcal{B}(p_2)$. Then:*

- $Q(X, Y) = \frac{1}{2}(p_1 - p_2 + 1)$, and
- X is statistically preferred to Y if and only if $p_1 \geq p_2$.

Proof: Let us compute the expression of the probabilistic relation $Q(X, Y)$:

$$\begin{aligned} Q(X, Y) &= P(X > Y) + \frac{1}{2}P(X = Y) \\ &= P(X = 1, Y = 0) + \frac{1}{2}\left(P(X = 0, Y = 0) + P(X = 1, Y = 1)\right) \\ &= p_1(1 - p_2) + \frac{1}{2}((1 - p_1)(1 - p_2) + p_1p_2) = \frac{1}{2}(p_1 - p_2 + 1). \end{aligned}$$

Then it holds that:

$$X \succeq_{\text{SP}} Y \Leftrightarrow Q(X, Y) \geq \frac{1}{2} \Leftrightarrow \frac{1}{2}(p_1 - p_2 + 1) \geq \frac{1}{2} \Leftrightarrow p_1 \geq p_2. \quad \blacksquare$$

Thus, a necessary and sufficient condition for $X \succeq_{\text{SP}} Y$ is that $p_1 \geq p_2$, or equivalently, $E[X] \geq E[Y]$. In fact, it is immediate that this condition is also necessary and sufficient for $X \succeq_{\text{FSD}} Y$. Thus, first degree stochastic dominance is a complete relation for Bernoulli distributions; as a consequence, the same applies to n -th degree stochastic dominance, and therefore they are equivalent methods. This allows us to establish the following corollary.

Corollary 3.66 *Let X and Y be two independent random variables with Bernoulli distribution. Then:*

$$X \succeq_{\text{FSD}} Y \Leftrightarrow X \succeq_{\text{nSD}} Y \text{ for any } n \geq 2 \Leftrightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

Continuous distributions under independence

Next, we consider some of the most important families of continuous distributions: exponential, beta, Pareto and uniform. In addition, due to the importance of the normal distribution, we devote the next paragraph to its study; in that case we shall consider other possibilities in addition to independent random variables.

Remark 3.67 *Although the beta distribution depends on two parameters, $p, q > 0$, in this work we shall consider the particular cases where one of the parameters equals 1, as in [56]. The general case in which both parameters are greater than 1 is much more complex, since the expression of the probabilistic relation is very difficult to obtain.*

Analogously, the Pareto distribution depends on two parameters, a, b , and the density function is given by

$$f(x) = \frac{ab^a}{x^{a+1}}, \quad x > b.$$

As in [56] we will focus on the case $b = 1$. ♦

Before starting, we recall in Table 3.3 the density functions and the parameters of the distributions we study along this subsection.

Distribution	Density function	Parameters
Exponential	$\lambda e^{-\lambda x}, x \in (0, \infty)$	$\lambda > 0$
Uniform	$\frac{1}{b-a}, x \in (a, b)$	$a, b \in \mathbb{R}, a < b$
Pareto	$\lambda x^{-(\lambda+1)}, x \in (1, \infty)$	$\lambda > 0$
Beta	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, x \in (0, 1)$	$p, q > 0$

Table 3.3: Characteristics of the continuous distributions to be studied.

Proposition 3.68 *Let X and Y be two independent random variables with exponential distributions, $X \equiv \text{Exp}(\lambda_1)$ and $Y \equiv \text{Exp}(\lambda_2)$, respectively. Then:*

- $Q(X, Y) = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ and
- X is statistically preferred to Y if and only if $\lambda_1 \leq \lambda_2$.

Proof: We first prove that $Q(X, Y) = \frac{\lambda_2}{\lambda_1 + \lambda_2}$.

$$\begin{aligned} Q(X, Y) &= P(X > Y) = \int_0^\infty \lambda_1 e^{-\lambda_1 x} dx \int_0^x \lambda_2 e^{-\lambda_2 y} dy = \int_0^\infty \lambda_1 e^{-\lambda_1 x} (1 - e^{-\lambda_2 x}) dx \\ &= \int_0^\infty \lambda_1 e^{-\lambda_1 x} dx - \int_0^\infty \lambda_1 e^{-(\lambda_1 + \lambda_2)x} dx = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned}$$

Thus,

$$X \succeq_{\text{SP}} Y \Leftrightarrow Q(X, Y) \geq \frac{1}{2} \Leftrightarrow \frac{\lambda_2}{\lambda_1 + \lambda_2} \geq \frac{1}{2} \Leftrightarrow \lambda_2 \geq \lambda_1. \quad \blacksquare$$

Remark 3.69 *The value of the probabilistic relation Q for independent and exponentially distributed random variables was already studied in [56, Section 6.2.1]. However, in such reference the authors made a mistake during the computations and found an incorrect expression for the probabilistic relation.♦*

As with Bernoulli distributed random variables, statistical preference and stochastic dominance are equivalent properties for exponential distributions. In this case, also first degree stochastic dominance, and therefore the n -degree stochastic dominance, are complete relations, and the can be reduced to the comparison of the expectations.

Corollary 3.70 *Let X and Y be two independent random variables with exponential distribution. Then,*

$$X \succeq_{\text{FSD}} Y \Leftrightarrow X \succeq_{\text{nSD}} Y \text{ for any } n \geq 2 \Leftrightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

Next we focus on uniform distributions.

Proposition 3.71 *Let X and Y be two independent random variables with uniform distributions, $U(a, b)$ and $U(c, d)$ respectively.*

1. *If $(a, b) \subseteq (c, d)$ then:*

- $Q(X, Y) = \frac{2b-c-d}{2(b-a)}$ and
- $X \succeq_{\text{SP}} Y$ if and only if $a + b \geq c + d$.

2. *If $c \leq a < d \leq b$, X is always statistically preferred to Y , and its degree of preference is $Q(X, Y) = 1 - \frac{(d-a)^2}{2(b-a)(d-c)}$.*

Proof:

1. Suppose that $a \leq c < d \leq b$. Then,

$$\begin{aligned} Q(X, Y) &= P(X > Y) = \int_d^b \frac{1}{b-a} dx + \int_c^d \int_c^x \frac{1}{b-a} \frac{1}{d-c} dy dx \\ &= \frac{b-d}{b-a} + \int_c^d \frac{1}{b-a} \frac{x-c}{d-c} dx = \frac{b-d}{b-a} + \frac{(d-c)^2}{2(d-c)(b-a)} = \frac{2b-c-d}{2(b-a)}. \end{aligned}$$

Then, $X \succeq_{\text{SP}} Y$ if and only if:

$$\frac{2b-c-d}{2(b-a)} \geq \frac{1}{2} \Leftrightarrow b+a \geq c+d.$$

If $c \leq a < b \leq d$, we can similarly see that

$$Q(X, Y) = \frac{b+a-2c}{2(d-c)}.$$

Thus, $Q(X, Y) \geq \frac{1}{2}$ if and only if $a + b \geq c + d$.

2. If $c \leq a < d \leq b$, it is easy to prove that $X \succeq_{\text{FSD}} Y$, and therefore $X \succeq_{\text{SP}} Y$. Let us now compute the preference degree:

$$P(Y > X) = \int_a^d \int_a^y \frac{dx dy}{(b-a)(d-c)} = \int_a^d \frac{y-a}{(b-a)(d-c)} dy = \frac{(d-a)^2}{2(b-a)(d-c)}.$$

Then, $Q(X, Y) = 1 - Q(Y, X) = 1 - P(Y > X) = 1 - \frac{(d-a)^2}{2(b-a)(d-c)}$. ■

Remark 3.72 *The value of the probabilistic relation Q for the uniform distribution was already studied in [56]. However, the authors only focused on uniform distribution with a fixed amplitude of the support, and the only parameter was the starting point of the support. This is a particular case included in the last result, and in that case, as we have seen, the random variable with the greatest minimum of the support stochastically dominates the other one, and consequently it is also statistically preferred. \blacklozenge*

For uniform distributions, first degree stochastic dominance and statistical preference are not equivalent in general. In fact, first degree stochastic dominance does not hold when the first case of the proof of the previous proposition holds. Nevertheless, we can establish the following:

Corollary 3.73 *Let X and Y be two independent random variables with uniform distribution. It holds that:*

$$X \succeq_{\text{FSD}} Y \Rightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

We next focus on the family of Pareto distribution.

Proposition 3.74 *Let X and Y be two independent random variables with Pareto distributions, $X \equiv \mathcal{P}_a(\lambda_1)$ and $Y \equiv \mathcal{P}_a(\lambda_2)$, respectively. Then:*

- $Q(X, Y) = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ and
- X is statistically preferred to Y if and only if $\lambda_2 \geq \lambda_1$.

Proof: First of all, let us determine the expression of Q :

$$\begin{aligned} Q(X, Y) &= P(X > Y) = \int_1^\infty \int_1^x \lambda_1 x^{-\lambda_1-1} \lambda_2 y^{-\lambda_2-1} dy dx \\ &= \int_1^\infty \lambda_1 x^{-\lambda_1-1} (1 - x^{-\lambda_2}) dx = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

Then,

$$X \succeq_{\text{SP}} Y \Leftrightarrow 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \geq \frac{1}{2} \Leftrightarrow \lambda_2 \geq \lambda_1. \quad \blacksquare$$

As for exponential and Bernoulli distributions, the equivalence between first degree stochastic dominance and statistical preference holds for Pareto distributions. In fact, when the expectation of the random variables exists, first degree stochastic dominance is equivalent to the comparison of the expectations. Hence, it is a complete relation, and then n -th degree stochastic dominance is also complete and equivalent to first degree stochastic dominance.

Corollary 3.75 *Let X and Y be two independent random variables with Pareto distributions. Then:*

$$X \succeq_{\text{FSD}} Y \Leftrightarrow X \succeq_{\text{nSD}} Y \text{ for any } n \geq 2 \Leftrightarrow X \succeq_{\text{SP}} Y.$$

Furthermore, if the parameter of X and Y are greater than 1, their expectation exists, and in that case:

$$X \succeq_{\text{FSD}} Y \Leftrightarrow X \succeq_{\text{nSD}} Y \text{ for any } n \geq 2 \Leftrightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

Concerning the beta distribution, we recall that its density function is given by

$$f(x) = \begin{cases} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (3.17)$$

However, the results we investigate in this section fix the value of one of the parameters to 1. We start by fixing $q = 1$. We obtain the following:

Proposition 3.76 *Let X and Y be two independent random variables with beta distributions, $X \equiv \beta(p_1, 1)$ and $Y \equiv \beta(p_2, 1)$, respectively. Then:*

- $Q(X, Y) = \frac{p_1}{p_1 + p_2}$ and
- $X \succeq_{\text{SP}} Y$ if and only if $p_1 \geq p_2$.

Proof: We first compute the expression of the relation Q .

$$Q(X, Y) = P(X > Y) = \int_0^1 \int_0^x p_1 x^{p_1-1} p_2 y^{p_2-1} dy dx = \int_0^1 p_1 x^{p_1-1} x^{p_2} dx = \frac{p_1}{p_1 + p_2}.$$

Then it holds that

$$X \succeq_{\text{SP}} Y \Leftrightarrow \frac{p_1}{p_1 + p_2} \geq \frac{1}{2} \Leftrightarrow p_1 \geq p_2. \quad \blacksquare$$

Taking into account that the expectation of a beta distribution with parameter $q = 1$ is $\frac{p}{p+1}$, the equivalence between statistical preference and the comparison of the expectations is clear. Furthermore, take into account that the cumulative distribution function associated with a beta distribution with parameter $q = 1$ is given by:

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0. \\ x^p & \text{if } 0 < x < 1. \\ 1 & \text{if } x \geq 1. \end{cases}$$

Then, it is clear that stochastic dominance between two variables of this type can be reduced to verifying which of the parameters p is greater. Finally, it is easy to check that this is equivalent to take the variable with greater expectation. Thus, in this case stochastic dominance, statistical preference and the comparison of expectations are also equivalent.

Corollary 3.77 *Let X and Y be two independent random variables with beta distributions with second parameter equal to 1. Then,*

$$X \succeq_{\text{FSD}} Y \Leftrightarrow X \succeq_{\text{nSD}} Y \text{ for any } n \geq 2 \Leftrightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

Finally, we consider beta distributions with $p = 1$.

Proposition 3.78 *Let X and Y be two independent random variables with distributions $X \equiv \beta(1, q_1)$ and $Y \equiv \beta(1, q_2)$, respectively. Then:*

- $Q(X, Y) = \frac{q_2}{q_1 + q_2}$ and
- $X \succeq_{\text{SP}} Y$ if and only if $q_2 \geq q_1$.

Proof: In order to prove the result, note that $X \equiv \beta(1, q) \Leftrightarrow 1 - X \equiv \beta(q, 1)$:

$$F_{1-X}(t) = P(1 - X \leq t) = 1 - F_X(1 - t) = 1 - [1 - (1 - (1 - t))^q] = t^q.$$

Then, taking into account Proposition 3.3, $X \succeq_{\text{SP}} Y \Leftrightarrow 1 - Y \succeq_{\text{SP}} 1 - X$ and $Q(X, Y) = Q(1 - Y, 1 - X) = \frac{q_2}{q_1 + q_2}$, and using Proposition 3.76, statistical preference is equivalent to $q_2 \geq q_1$. ■

As in the previous case, since the expectation of a beta distribution with parameter $p = 1$ is $\frac{1}{1+q}$, the equivalence between stochastic dominance and statistical preference also holds for beta distributions.

Corollary 3.79 *Let X and Y be two independent random variables with beta distributions with first parameter equal to 1. Then,*

$$X \succeq_{\text{FSD}} Y \Leftrightarrow X \succeq_{\text{nSD}} Y \text{ for any } n \geq 2 \Leftrightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

The normal distribution

We now study normally distributed random variables. In this case we will not only consider independent variables. Thus, we begin with the comparison of one-dimensional distributions and then we shall consider the case of the comparison of the components of a bidimensional random vector normally distributed.

Proposition 3.80 *Let X and Y be two independent and normally distributed random variables, $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, respectively. Then, X will be statistically preferred to Y if and only if $\mu_1 \geq \mu_2$.*

Proof: The relation Q takes the value (see [56, Section 7]):

$$Q(X, Y) = F_{N(0,1)} \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right).$$

Then:

$$X \succeq_{\text{SP}} Y \Leftrightarrow Q(X, Y) \geq \frac{1}{2} \Leftrightarrow F_{N(0,1)} \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right) \geq \frac{1}{2} \Leftrightarrow \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \geq 0 \Leftrightarrow \mu_1 \geq \mu_2. \quad \blacksquare$$

Given two normally distributed random variables $X \sim N(\mu_1, \sigma_1)$ and $Y \sim N(\mu_2, \sigma_2)$, it holds that $X \succeq_{\text{FSD}} Y$ if and only if they are identically distributed, $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, (see [139]). Then, statistical preference is not equivalent to first degree stochastic dominance for normal random variables.

For independent normal distributions, the variance of the variables are not important when studying statistical preference. For this reason, statistical preference is equivalent to the criterium of maximum mean in the comparison of normal random variables:

Corollary 3.81 *Consider two independent random variables X and Y normally distributed. It holds that:*

$$X \succeq_{\text{FSD}} Y \Rightarrow X \succeq_{\text{SP}} Y \Leftrightarrow E[X] \geq E[Y].$$

Let us now consider a bidimensional random vector with normal distribution:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix} \right). \quad (3.18)$$

Now, our aim is to compare the components X_1 and X_2 of this random vector. We obtain the following result:

Theorem 3.82 *Consider the random vector $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ normally distributed as in Equation (3.18). Then, it holds that:*

- $Q(X_1, X_2) = F_{N(0,1)} \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right).$

- $X_1 \succeq_{\text{SP}} X_2 \Leftrightarrow \mu_1 \geq \mu_2$.

Proof: Applying the usual properties of the normal distributions, the distribution of $X_1 - X_2$ is:

$$\begin{aligned} X_1 - X_2 &= (1 \quad -1) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv N \left((1 \quad -1) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, (1 \quad -1) \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) \\ &= N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2), \end{aligned}$$

where the second parameter is consider to be the variance instead of the standard deviation. Then:

$$\begin{aligned} P(X_1 > X_2) &= P(X_1 - X_2 > 0) = P \left(N(0, 1) > \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right) \\ &= P \left(N(0, 1) < \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right) = F_{N(0,1)} \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right). \end{aligned}$$

Thus, $X_1 \succeq_{\text{SP}} X_2$ if and only if $F_{N(0,1)} \left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right) \geq \frac{1}{2}$. ■

This result is more general than Proposition 3.80, which corresponds to the case $\rho = 0$. Moreover, in that case statistical preference is also equivalent to the comparison of the expectations. However, the advantage of obtaining a degree of preference is obvious. In fact, we have to recall the influence of the correlation coefficient ρ in the value of the preference degree: although the preference between X_1 and X_2 is only based on the comparison of the expectations ($X_1 \succeq_{\text{SP}} X_2 \Leftrightarrow \mu_1 \geq \mu_2$), the value of ρ plays an important role for the preference degree. For instance, the greater the correlation coefficient, the greater the preference degree $Q(X, Y)$. For this reason, the greater the correlation coefficient, the stronger the preference of X over Y .

In Table 3.4 we have summarised the results that we have obtained in this subsection.

As a summary, we have seen that for the some of usual distributions in independent random variables, statistical preference is equivalent to the comparison of its expectations, and in several cases, stochastic dominance and statistical preference are also equivalent. Let us recall that, in particular, for the distributions we have studied that belongs to the exponential family of distributions, stochastic dominance and statistical preference are equivalent. We can conjecture that for independent random variables whose distribution belong to the exponential family of distributions, statistical preference and stochastic dominance are equivalent, and are also equivalent to the comparison of the expectations.

Nevertheless, at this point, this is just a conjecture because it has not been proved yet.

Distributions	$Q(X_1, X_2)$	Condition
$X_i \equiv B(p_i), i = 1, 2$	$\frac{1}{2}(p_1 - p_2 + 1)$	$p_1 \geq p_2$
$X_i \equiv Exp(\lambda_i), i = 1, 2$	$\frac{\lambda_2}{\lambda_1 + \lambda_2}$	$\lambda_2 \geq \lambda_1$
$X_1 \equiv U(a, b), X_2 \equiv U(c, d)$		
$a \leq c \leq d < b$	$\frac{2b-c-d}{2(b-a)}$	$a + b \geq c + d$
$c < a < b \leq d$	$\frac{a+b-2c}{2(d-c)}$	$a + b \geq c + d$
$c \leq a < d \leq b$	$1 - \frac{(d-a)^2}{2(d-c)(b-a)}$	Always
$\mathcal{P}_a(\lambda_i), i = 1, 2$	$\frac{\lambda_2}{\lambda_1 + \lambda_2}$	$\lambda_2 \geq \lambda_1$
$\beta(p_i, 1), i = 1, 2$	$\frac{p_1}{p_1 + p_2}$	$p_1 \geq p_2$
$\beta(1, q_i), i = 1, 2$	$\frac{q_2}{q_1 + q_2}$	$q_2 \geq q_1$
$N(\mu_i, \sigma_i), i = 1, 2$	$F_{N(0,1)}\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$	$\mu_1 \geq \mu_2$

Table 3.4: Characterizations of statistical preference between independent random variables included in the same family of distributions.

Although during this paragraph we have assumed independence for non-normally distributed variables, there are other cases of interest. For instance, in [32] the case of comonotonic and countermonotonic random variables are studied. In particular, Proposition 3.65, that assures that

$$X \succeq_{SP} Y \Leftrightarrow X \succeq_{nSD} Y \Leftrightarrow E[X] \geq E[Y] \text{ for any } n \geq 1$$

for independent random variables with Bernoulli distribution, could be easily extended to Bernoulli distributed random variables, taking into account the possible dependence relationship between them.

3.3 Comparison of n variables by means of the statistical preference

So far, we have investigated several properties of stochastic dominance and statistical preference as pairwise comparison methods. However, a natural question arises: can

we employ those methods for the comparison of more than two variables? On the one hand, stochastic dominance was defined as a pairwise comparison method, based on the direct comparison of the cumulative distribution functions, or their iterative integrals. As we already mentioned, stochastic dominance allows for incomparability. Thus, if incomparability can happen when comparing two distribution functions, it should be more frequent when comparing more than two. Then, stochastic dominance does not seem to be a good alternative for the comparison of more than two variables.

On the other hand, statistical preference has an important drawback: its lack of transitivity. The idea of statistical preference is to consider X preferred to Y when it provides greater utility the majority of times. As such, it is close to the rule of majority in voting systems; taking into account Condorcet's paradox (see [40]) it is not difficult to see that statistical preference is not transitive. When De Schuymer et al. ([55, 57]) introduced this notion, they provided an example to illustrate this fact; another one can be found in [67, Example 3].

Example 3.83 ([57, Section 1]) *As in Example 3.10, consider the following dice:*

$$\begin{aligned} A &= \{1, 3, 4, 15, 16, 17\}, \\ B &= \{2, 10, 11, 12, 13, 14\}, \end{aligned} \tag{3.19}$$

and also the dice

$$C = \{5, 6, 7, 8, 9, 18\},$$

where by dice we mean a discrete and uniformly distributed random variable. We consider the game consisting on rolling the three dice simultaneously, so that the dice whose number is greater wins the game. Thus, A , B and C can be seen as independent random variables.

If we compute the probabilistic relation Q for these dices we obtain the following results:

$$\begin{aligned} Q(A, B) &= \frac{5}{9} \Rightarrow A \succ_{\text{SP}} B. \\ Q(B, C) &= \frac{25}{36} \Rightarrow B \succ_{\text{SP}} C. \\ Q(C, A) &= \frac{7}{12} \Rightarrow C \succ_{\text{SP}} A. \end{aligned}$$

Hence, dice A is statistically preferred to dice B , dice B is statistically preferred to dice C but dice C is statistically preferred to dice A , that is, there is a cycle, as we can see in Figure 3.4.♦

This fact is known as the cycle-transitivity problem, and it has already been studied by some authors, like De Shuymer et al. ([14, 15, 16, 49, 54, 56, 57, 58]) and Martinetti et al. ([122]).

This shows that statistical preference could not be adequate when we want to compare more than two random variables, precisely because it is based on pairwise comparisons.

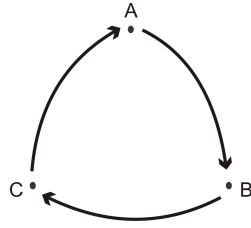


Figure 3.4: Probabilistic relation for the three dices.

Since both stochastic dominance and statistical preference do not seem to be adequate methods for the comparison of more than two variables, our aim in this section is to provide a generalisation of the statistical preference for the comparison of n random variables, based on an extension of the probabilistic relation defined in Equation (2.7). After introducing the main definition, we shall investigate its properties, its possible characterizations and its connection with the “usual” statistical preference, as well as its possible relationships with stochastic dominance.

3.3.1 generalisation of the statistical preference

First of all we are going to analyze the case of three random variables, as in the dice example, and later we shall generalise our definition to the case of n random variables.

Let us consider three random variables denoted by X , Y and Z defined on the probability space (Ω, \mathcal{A}, P) . We can decompose Ω in the following way:

$$\begin{aligned} \Omega = & \{X > \max(Y, Z)\} \cup \{Y > \max(X, Z)\} \cup \{Z > \max(X, Y)\} \\ & \cup \{X = Y > Z\} \cup \{X = Z > Y\} \cup \{Y = Z > X\} \cup \{X = Y = Z\} \end{aligned} \quad (3.20)$$

Obviously, $\{X > \max(Y, Z)\}$ denotes the subset of Ω formed by the elements $\omega \in \Omega$ satisfying $X(\omega) > \max(Y(\omega), Z(\omega))$, and similarly for the others. In what remains we will use the short way in order to simplify the notation.

This is a decomposition of Ω into pairwise disjoint subsets, i.e., a partition of Ω . As a consequence,

$$\begin{aligned} 1 = & P(X > \max(Y, Z)) + P(Y > \max(X, Z)) + P(Z > \max(X, Y)) + P(X = Y > Z) \\ & + P(X = Z > Y) + P(Y = Z > X) + P(X = Y = Z). \end{aligned} \quad (3.21)$$

Since our goal is to define the degree in which X is preferred to Y and Z , we can define $Q_2(X, [Y, Z])$ by the following equation:

$$Q_2(X, [Y, Z]) = P(X > \max(Y, Z)) + \frac{1}{2}(P(X = Y > Z) + P(X = Z > Y)) + \frac{1}{3}P(X = Y = Z).$$

This generalises Equation (2.7). Furthermore, if we consider $Q_2(Y, [X, Z])$ and $Q_2(Z, [X, Y])$, given by:

$$\begin{aligned} Q_2(Y, [X, Z]) &= P(Y > \max(X, Z)) + \frac{1}{2}(P(X = Y > Z) + P(Y = Z > X)) \\ &\quad + \frac{1}{3}P(X = Y = Z); \\ Q_2(Z, [X, Y]) &= P(Z > \max(X, Y)) + \frac{1}{2}(P(X = Z > Y) + P(Y = Z > X)) \\ &\quad + \frac{1}{3}P(X = Y = Z); \end{aligned}$$

using the partition of Ω showed in Equation (3.20) and Equation (3.21), it can be shown that:

$$Q_2(X, [Y, Z]) + Q_2(Y, [X, Z]) + Q_2(Z, [X, Y]) = 1.$$

In this sense, following the idea of De Schuymer et al. ([55, 57]), X can be considered preferred to Y and Z if

$$Q_2(X, [Y, Z]) \geq \max\{Q_2(Y, [X, Z]), Q_2(Z, [X, Y])\}.$$

Moreover, X is preferred to Y and Z with degree $Q_2(X, [Y, Z])$.

More generally, we can consider a set of alternatives \mathcal{D} formed by some random variables defined on the same probability space. Then, we can consider the map:

$$Q_n : \mathcal{D} \times \mathcal{D}^n \rightarrow [0, 1],$$

defined by:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= \text{Prob}\{X > \max(X_1, \dots, X_n)\} \\ &\quad + \frac{1}{2} \sum_{i=1}^n \text{Prob}\{X = X_i > \max(X_j : j \neq i)\} \\ &\quad + \frac{1}{3} \sum_{1 \leq i < j \leq n} \text{Prob}\{X = X_i = X_j > \max(X_k : k \neq i, j)\} \\ &\quad + \dots + \frac{1}{n+1} \text{Prob}\{X = X_1 = \dots = X_n\}. \end{aligned}$$

Equivalently, the relation Q_n can be expressed by:

$$Q_n(X, [X_1, \dots, X_n]) = \sum_{\substack{k=0, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n}} \frac{1}{k+1} P(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i_1, \dots, i_k} (X_j)), \quad (3.22)$$

where $\{i_1, \dots, i_k\}$ denotes any ordered subset of k -elements of $\{1, \dots, n\}$. Note that this formula is the generalisation of the probabilistic relation defined on Equation (2.7), since for $n = 1$ we obtain the expression of such probabilistic relation. We can interpret the value of $Q_n(X, [X_1, \dots, X_n])$ as the degree in which X is preferred to X_1, \dots, X_n . Consequently, the greater the value of $Q_n(X, [X_1, \dots, X_n])$ the stronger the preference of X over X_1, \dots, X_n . The relation Q_n allows to define the concept of general statistical preference.

Definition 3.84 *Let X, X_1, \dots, X_n be $n+1$ random variables. X is statistically preferred to X_1, \dots, X_n , and it is denoted by $X \succeq_{\text{SP}} [X_1, \dots, X_n]$, if*

$$Q_n(X, [X_1, \dots, X_n]) \geq \max_{i=1, \dots, n} Q_n(X_i, [X, \{X_j : j \neq i\}]). \quad (3.23)$$

As it was the case for statistical preference, this generalisation uses the joint distribution of the variables, and thus takes into account the stochastic dependencies between them. Moreover, the relation Q_n provides a degree of preference of a random variable with respect to the others, and through this we can establish which is the preferred random variable, the second preferred random variable, etc. For instance, if $Q_n(X_i, [X, \{X_j : j \neq i\}]) \geq Q_n(X_j, [X, \{X_j : j \neq i\}])$ for every $i > j$ and Equation (3.23) holds, then X is the preferred random variable, X_1 is the second preferred random variable and, in general, X_i is the $i + 1$ preferred random variable, with their respective degrees of preference.

Example 3.85 *If we consider the dices defined on Equation (3.19) and apply the general statistical preference to find the preferred dice, we obtain the following preference degrees: $Q_2(X, [Y, Z]) = 0.4167$; $Q_2(Y, [X, Z]) = 0.3472$; and $Q_2(Z, [X, Y]) = 0.2361$. Thus, X is the preferred dice with degree 0.4167; Y is the second preferred dice with degree 0.3472; and Z is the less preferred dice with degree 0.2361. ♦*

3.3.2 Basic properties

In this subsection we investigate some basic properties of the general statistical preference. The first part is devoted to the study of the relationships between pairwise statistical preference and general preference. Similarly, we also establish a connection between $Q(\cdot, \cdot)$ and $Q_n(\cdot, [\cdot])$. Finally, we generalise Proposition 3.39 and Theorem 3.40, where we showed the connection between statistical preference and the median, for the general statistical preference and establish a characterization of this notion.

Consider random variables X, X_1, \dots, X_n . In our first result we prove that general statistical preference sometimes offers a different preferred random variable than pairwise statistical preference. This is because general statistical preference uses the joint distribution of all the variables, while pairwise statistical preference only takes into account their bivariate distributions, and consequently it does not use all the available information.

Proposition 3.86 *Let X, X_1, \dots, X_n be $n + 1$ random variables. It holds that:*

- *There are X, X_1, \dots, X_n random variables such that $X \succeq_{\text{SP}} X_i$ for every $i = 1, \dots, n$ and $X_j \succeq_{\text{SP}} [X, X_i : i \neq j]$ for some $j \in \{1, \dots, n\}$.*
- *There are X, X_1, \dots, X_n random variables such that $X_i \succeq_{\text{SP}} X$ for every $i = 1, \dots, n$ and $X \succeq_{\text{SP}} [X_1, \dots, X_n]$.♦*

Proof: Let us consider the first statement. To see that the implication does not hold in general, consider $n = 2$ and the independent random variables X, X_1 and X_2 defined by:

$$\begin{array}{c|cc} X & 3 & 5 \\ \hline P_X & 0.5 & 0.5 \end{array} \quad \begin{array}{c|cc} X_1 & 0 & 5 \\ \hline P_{X_1} & 0.5 & 0.5 \end{array} \quad \begin{array}{c|cc} X_2 & 2 & 6 \\ \hline P_{X_2} & 0.51 & 0.49 \end{array}$$

For these variables it holds that $Q(X, X_1) = 0.625$ and $Q(X, X_2) = 0.51$, and consequently $X \succeq_{\text{SP}} X_1$ and $X \succeq_{\text{SP}} X_2$. However,

$$\begin{aligned} Q_2(X, [X_1, X_2]) &= 0.31875. \\ Q_2(X_1, [X, X_2]) &= 0.19125. \\ Q_2(X_2, [X, X_1]) &= 0.49. \end{aligned}$$

Thus, $X_2 \succeq_{\text{SP}} [X, X_1]$.

Consider now the second statement. Consider $n = 2$ and the independent dices X, X_1 and X_2 defined by:

$$\begin{aligned} X &= \{1, 2, 4, 6, 17, 18\}. \\ X_1 &= \{3, 7, 9, 12, 14, 16\}. \\ X_2 &= \{5, 8, 10, 11, 13, 15\}. \end{aligned}$$

It holds that $X_1 \succeq_{\text{SP}} X$ and $X_2 \succeq_{\text{SP}} X$, since $Q(X, X_1) = \frac{7}{18}$ and $Q(X, X_2) = \frac{13}{36}$. However, if we compute the relation $Q_2(\cdot, [\cdot])$ we obtain the following:

$$\begin{aligned} Q_2(X, [X_1, X_2]) &= \frac{73}{216}. \\ Q_2(X_1, [X, X_2]) &= \frac{72}{216}. \\ Q_2(X_2, [X, X_1]) &= \frac{71}{216}. \end{aligned}$$

Consequently, $X \succeq_{\text{SP}} [X_1, X_2]$. ■

Next we prove that $Q_n(X, [X_1, \dots, X_n])$ is always lower than or equal to $Q(X, X_i)$.

Proposition 3.87 *Let us consider the random variables X, X_1, \dots, X_n . It holds that:*

$$Q_n(X, [X_1, \dots, X_n]) \leq Q(X, X_i) \text{ for every } i = 1, \dots, n.$$

Consequently, if $Q_n(X, [X_1, \dots, X_n]) \geq \frac{1}{2}$, then $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ and $X \succeq_{\text{SP}} X_i$ for every $i = 1, \dots, n$.

Proof: Recall that $Q(X, X_i) = P(X > X_i) + \frac{1}{2}P(X = X_i)$. It holds that:

$$\{X > X_i\} \supseteq \bigcup_{\substack{k=0, \dots, n-1 \\ i_1, \dots, i_k \neq i}} \left\{ X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, i_1, \dots, i_k} (X_i, X_j) \right\}.$$

Moreover, the previous sets are pairwise disjoint, and consequently:

$$P(X > X_i) \geq \sum_{\substack{k=0, \dots, n-1 \\ i_1, \dots, i_k \neq i}} P\left(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, i_1, \dots, i_k} (X_i, X_j)\right).$$

Similarly:

$$\{X = X_i\} \supseteq \bigcup_{\substack{k=0, \dots, n-1 \\ i_1, \dots, i_k \neq i}} \left\{ X = X_i = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, i_1, \dots, i_k} (X_j) \right\}.$$

Since these sets are pairwise disjoint,

$$P(X = X_i) \geq \sum_{\substack{k=0, \dots, n-1 \\ i_1, \dots, i_k \neq i}} P\left(X = X_i = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, \dots, i_k} (X_j)\right).$$

Consequently, we obtain that:

$$\begin{aligned} Q(X, X_i) &= P(X > X_i) + \frac{1}{2}P(X = X_i) \geq \\ &\sum_{\substack{k=0, \dots, n-1 \\ i_1, \dots, i_k \neq i}} P(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, i_1, \dots, i_k} (X_i, X_j)) + \\ &\frac{1}{2} \sum_{\substack{k=0, \dots, n-1 \\ i_1, \dots, i_k \neq i}} P(X = X_i = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, \dots, i_k} (X_j)) \geq \\ &\sum_{\substack{k=0, \dots, n \\ i_1, \dots, i_k \in \{1, \dots, n\}}} \frac{1}{k+1} P(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i_1, \dots, i_k} (X_j)) = \\ &Q_n(X, [X_1, \dots, X_n]). \end{aligned}$$

We conclude that $Q(X, X_i) \geq Q_n(X, [X_1, \dots, X_n])$. Consequently, if

$$Q_n(X, [X_1, \dots, X_n]) \geq \frac{1}{2}$$

then $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ and $X \succeq_{\text{SP}} X_i$ for every $i = 1, \dots, n$. ■

Next we establish the connection between the probabilistic relation $Q(\cdot, \cdot)$ and $Q_n(\cdot, [\cdot])$.

Proposition 3.88 *Let X, X_1, \dots, X_n be $n + 1$ random variables defined on the same probability space. It holds that*

$$Q_n(X, [X_1, \dots, X_n]) - Q(X, \max(X_1, \dots, X_n)) = \sum_{k=2}^n \left(\frac{1}{k+1} - \frac{1}{2} \right) \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n \\ i_j \neq i_l \forall j \neq l}} P(X = X_{i_1} = \dots = X_{i_k} > \max_{l \neq i_1, \dots, i_k} (X_l)).$$

Proof: Consider the expression of $Q(X, \max(X_1, \dots, X_n))$:

$$Q(X, \max(X_1, \dots, X_n)) = P(X > \max(X_1, \dots, X_n)) + \frac{1}{2} \left(\sum_{k=1}^n \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n \\ i_j \neq i_l \forall j \neq l}} P(X = X_{i_1} = \dots = X_{i_k} > \max_{l \neq i_1, \dots, i_k} (X_l)) \right).$$

Using Equation (3.22), $Q_n(X, [X_1, \dots, X_n])$ can be expressed by:

$$Q_n(X, [X_1, \dots, X_n]) = P(X > \max(X_1, \dots, X_n)) + \sum_{k=1}^n \frac{1}{k+1} \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n \\ i_j \neq i_l \forall j \neq l}} P(X = X_{i_1} = \dots = X_{i_k} > \max_{l \neq i_1, \dots, i_k} (X_l)).$$

The result follows simply by making the difference between both expressions. ■

From this result we deduce that

$$Q_n(X, [X_1, \dots, X_n]) \leq Q(X, \max(X_1, \dots, X_n)). \quad (3.24)$$

Then, if $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ holds with degree $Q_n(X, [X_1, \dots, X_n]) \geq \frac{1}{2}$, we obtain $X \succeq_{\text{SP}} \max(X_1, \dots, X_n)$.

Moreover, there are situations where the inequality of Equation (3.24) becomes an equality. To see this, let us introduce the following notation:

$$X_{-i} = \{X_j : j \neq i\}.$$

Corollary 3.89 *Under the conditions of the previous proposition, if for every $k \in \{1, \dots, n\}$ and for every $1 \leq i_1 < \dots < i_k$ it holds that*

$$P(X = X_{i_1} = \dots = X_{i_k} > \max(X_j : j \neq i_1, \dots, i_k)) = 0, \quad (3.25)$$

then

$$Q_n(X, [X_1, \dots, X_n]) = Q(X, \max(X_1, \dots, X_n)).$$

Furthermore, if for every $k \in \{1, \dots, n\}$ and for every $1 \leq i_1 < \dots < i_k \leq n$ it holds that

$$P(X_{i_1} = \dots = X_{i_k} > \max(X, X_j : j \neq i_1, \dots, i_k)) = 0, \quad (3.26)$$

then

$$Q_n(X_i, [X, X_{-i}]) = Q(X_i, \max(X, X_{-i})),$$

for every $i = 1, \dots, n$.

In particular, the previous result holds when the random variables satisfy, $P(X = X_i) = P(X = X_j) = 0$ for every $i \neq j$, as is for instance the case with discrete random variables with pairwise disjoint supports.

Finally, let us generalise Theorem 3.40 and to provide a characterization of general statistical preference. For this aim we consider random variables X, X_1, \dots, X_n satisfying Equations (3.25) and (3.26) for every $k \in \{0, \dots, n\}$ and every $1 \leq i_1 < \dots < i_k \leq n$. Although this restriction will be imposed also in Theorems 3.91, 3.95 and Lemma 3.94, it is not too restrictive. In fact, it is satisfied by discrete random variables with pairwise disjoint supports or absolutely continuous random vectors (X, X_1, \dots, X_n) . Consequently, we can understand it as a technical condition.

Theorem 3.90 *Let X, X_1, \dots, X_n be $n + 1$ real-valued random variables defined on the same probability satisfying Equations (3.25) and (3.26). Then, $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ holds if and only if*

$$F_{X - \max(X_1, \dots, X_n)}(0) \leq F_{X_i - \max(X, X_{-i})}(0) \text{ for every } i = 1, \dots, n.$$

Proof: The probabilistic relation $Q(X, Y)$ can be expressed by:

$$Q(X, Y) = 1 - F_{X - Y}(0) + \frac{1}{2}P(X = Y).$$

Thus, using this expression and applying Corollary 3.89 it holds that:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= Q(X, \max(X_1, \dots, X_n)) = 1 - F_{X - \max(X_1, \dots, X_n)}(0) \\ &\quad + \frac{1}{2}P(X = \max(X_1, \dots, X_n)) = 1 - F_{X - \max(X_1, \dots, X_n)}(0). \end{aligned}$$

Similarly, we can compute the value of $Q_n(X_i, [X, X_{-i}])$:

$$Q_n(X_i, [X, X_{-i}]) = 1 - F_{X_i - \max(X, X_{-i})}(0).$$

Therefore, $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ if and only if:

$$1 - F_{X - \max(X_1, \dots, X_n)}(0) \geq 1 - F_{X_i - \max(X, X_{-i})}(0),$$

or equivalently,

$$F_{X-\max(X_1, \dots, X_n)}(0) \leq F_{X_i-\max(X, X_{-i})}(0)$$

for every $i = 1, \dots, n$. ■

Thus, given random variables X, X_1, \dots, X_n in the conditions of the previous result, to find the preferred one by computing the values of $Q_n(\cdot, [\cdot])$ is equivalent to comparing the values of $F_{X-\max(X_1, \dots, X_n)}(0)$ and $F_{X_i-\max(X, X_{-i})}(0)$ for $i = 1, \dots, n$.

3.3.3 Stochastic dominance Vs general statistical preference

In Section 3.2 we saw that in a number of cases first degree stochastic dominance implies statistical preference for real-valued random variables. Now we investigate the connection between stochastic dominance and general statistical preference. Again, we shall consider different cases: on the one hand, independent and comonotonic random variables, for which we shall obtain an equivalent expression for $Q_n(\cdot, [\cdot])$. On the other hand, we shall consider random variables coupled by Archimedean copulas. Recall that we omit countermonotonic random variables since, as we already said, the Łukasiewicz operator is not a copula for $n \geq 2$. Finally, we also investigate the relationships between the n^{th} degree stochastic dominance and general statistical preference.

Independent and comonotonic random variables

Let us begin our study with the case of independent real-valued random variables. In this case, by generalizing Theorem 3.44, we deduce that first degree stochastic dominance implies general statistical preference.

Theorem 3.91 *Let us consider X, X_1, \dots, X_n independent real-valued random variables satisfying Equations (3.25) and (3.26). Then, if $X \succeq_{\text{FSD}} X_i$ for $i = 1, \dots, n$, implies $X \succeq_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: Since we are under the hypotheses of Corollary 3.89, we deduce that:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= Q(X, \max(X_1, \dots, X_n)) \text{ and} \\ Q_n(X_i, [X, X_{-i}]) &= Q(X_i, \max(X, X_{-i})), \end{aligned}$$

for every $i = 1, \dots, n$. Therefore, $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ if and only if:

$$P(X \geq \max(X_1, \dots, X_n)) \geq P(X_i \geq \max(X, X_{-i})), \quad i = 1, \dots, n.$$

Note that, since X, X_1, \dots, X_n are independent, we also have that:

- X and $\max(X_1, \dots, X_n)$ are independent.

- X_i and $\max(X, X_{-i})$ are independent.

Now, we have to remark that, if U_1 and U_2 are two independent random variables with respective cumulative distribution functions F_{U_1} and F_{U_2} , Lemma 3.11 assures that $P\{U_1 \geq U_2\} = E[F_{U_2}(U_1)]$.

Applying this result, we deduce that:

$$P(X \geq \max(X_1, \dots, X_n)) = E(F_{\max(X_1, \dots, X_n)}(X)) = E(F_{X_1}(X) \dots F_{X_n}(X)).$$

Similarly,

$$\begin{aligned} P(X_i \geq \max(X, X_{-i})) &= E(F_{\max(X, X_{-i})}(X_i)) \\ &= E[F_X(X_i) \prod_{j \neq i} F_{X_j}(X_i)] \leq E[\prod_{j=1}^n F_{X_j}(X_i)], \end{aligned}$$

where last inequality holds since $F_X \leq F_{X_i}$. Finally, since $X \succeq_{\text{FSD}} X_i$, Equation (2.6) assures that $E[h(X)] \geq E[h(X_i)]$ for any increasing function h . In particular, we may consider the increasing function

$$h(t) = \prod_{j=1}^n F_{X_j}(t).$$

Therefore,

$$\begin{aligned} P(X \geq \max(X_1, \dots, X_n)) &= E(F_{X_1}(X) \dots F_{X_n}(X)) \\ &\geq E[\prod_{j=1}^n F_{X_j}(X)] \geq P(X_i \geq \max(X, X_{-i})), \end{aligned}$$

or equivalently,

$$Q(X, \max(X_1, \dots, X_n)) \geq Q(X_i, \max(X, X_{-i})).$$

We conclude that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. ■

Now we shall see that, as with statistical preference for independent random variables, strict first degree stochastic dominance also implies strict general statistical preference. For this aim, we need to establish the following lemma.

Lemma 3.92 *Consider $n + 1$ independent real-valued random variables X, X_1, \dots, X_n satisfying Equations (3.25) and (3.26) such that $X \succ_{\text{FSD}} X_i$ for $i = 1, \dots, n$. The following statements hold:*

1. *There is t^* such that $F_X(t^*) < F_{X_i}(t^*)$ and $F_{X_j}(t^*) > 0$ for any $j \neq i$.*
2. *If $P(X_i = t) = 0$ for any t satisfying the first point, then there exists an interval $[a, b]$, and $\varepsilon > 0$ such that:*

$$\prod_{j=1}^n F_{X_j}(t) - F_X(t) - \prod_{j \neq i} F_{X_j}(t) \geq \varepsilon > 0,$$

and $P(X_j \in [a, b]) > 0$.

Proof: Let us prove the first statement. Ex-absurdo, assume that for any t such that $F_X(t) < F_{X_i}(t)$, there exist j_1, \dots, j_k such that $F_{X_{j_1}}(t) = F_{X_{j_k}}(t) = 0 < F_{X_j}(t)$ for any $j \neq j_1, \dots, j_k$, and therefore $F_X(t) = 0$. Since the cumulative distribution functions are right-continuous, there is t^* such that $0 = F_X(t) < F_{X_i}(t)$ for any $t < t^*$ and $0 < F_X(t^*) \leq F_{X_j}(t^*)$ for any $j = 1, \dots, n$. Then:

$$P(X = t^*) > 0, P(X_{j_1} = t^*) > 0, \dots, P(X_{j_k} = t^*) > 0.$$

Hence:

$$P(X = X_{j_1} = \dots = X_{j_k} > X_j : j \neq j_1, \dots, j_k) \geq \\ P(X = X_{j_1} = \dots = X_{j_k} = t^* > X_j : j \neq j_1, \dots, j_k) > 0,$$

and this contradicts Equation (3.25). We conclude that there exists at least t^* such that $F_X(t^*) < F_{X_i}(t^*)$ and $F_{X_j}(t^*) > 0$ for any $j \neq i$.

Let us now check the second statement. Let t^* be a point such that $F_X(t^*) < F_{X_i}(t^*)$ and $F_{X_j}(t^*) > 0$ for any $j \neq i$. Following the same steps as in Lemma 3.45 we can prove that the existence of an interval $[a, b]$ including t^* and $\delta > 0$ such that $F_{X_i}(t) - F_X(t) \geq \delta > 0$ for any $t \in [a, b]$ and $P(X_i \in [a, b]) > 0$. Furthermore, since by hypothesis $P(X_i = t) = 0$ for any $t \in [a, b]$, F_{X_i} should be strictly increasing in a subinterval $[a_1, b_1]$ of $[a, b]$.

Now, consider a point t_0 in the interval $[a_1, b_1]$. Since all the F_{X_j} , for $j = 1, \dots, n$, and F_X are right-continuous:

$$\lim_{\varepsilon \rightarrow 0} \prod_{j=1}^n F_{X_j}(t_0 + \varepsilon) = \prod_{j=1}^n F_{X_j}(t_0) > F_X(t_0) \prod_{j \neq i} F_{X_j}(t_0) \\ = \lim_{\varepsilon \rightarrow 0} F_X(t_0 + \varepsilon) \prod_{j \neq i} F_{X_j}(t_0 + \varepsilon).$$

Then, there is $\varepsilon > 0$, and can we assume $\varepsilon \leq b_1 - t_0$, such that:

$$F_X(t_0 + \varepsilon) \prod_{j \neq i} F_{X_j}(t_0 + \varepsilon) \leq F_X(t_0) \prod_{j \neq i} F_{X_j}(t_0) + \frac{\prod_{j=1}^n F_{X_j}(t_0) - F_X(t_0) \prod_{j \neq i} F_{X_j}(t_0)}{2} \\ < \prod_{j=1}^n F_{X_j}(t_0).$$

Taking $\delta' = \frac{\prod_{j=1}^n F_{X_j}(t_0) - F_X(t_0) \prod_{j \neq i} F_{X_j}(t_0)}{2} > 0$, then:

$$\prod_{j=1}^n F_{X_j}(t) - F_X(t) \prod_{j \neq i} F_{X_j}(t) \geq \delta' > 0$$

for any $t \in [t_0, t_0 + \varepsilon]$. Moreover, since F_{X_i} is strictly increasing in $[a, b]$, it is also strictly increasing in $[t_0, t_0 + \varepsilon]$, and therefore $P(X_i \in [t_0, t_0 + \varepsilon]) > 0$. ■

Proposition 3.93 *Let X, X_1, \dots, X_n be $n+1$ independent real-valued random variables satisfying Equations (3.25) and (3.26). Then, if $X \succ_{\text{FSD}} X_i$ for any $i = 1, \dots, n$ it holds that $X \succ_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: Since $X \succ_{\text{FSD}} X_i$ implies $X \succeq_{\text{FSD}} X_i$, we know that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. Taking into account the previous result, it suffices to prove that $E[F_X(X_i) \prod_{j \neq i} F_{X_j}(X_i)] < E[\prod_{j=1}^n F_{X_j}(X_i)]$ for $i = 1, \dots, n$, since this implies that:

$$Q_n(X, [X_1, \dots, X_n]) \geq Q_n(X_i, [X, X_{-i}]) \text{ for } i = 1, \dots, n.$$

Using the previous lemma, we can assume there is t_0 such that $F_X(t_0) < F_{X_i}(t_0)$ and $F_{X_j}(t_0) > 0$ for any $j \neq i$.

Consider two cases:

- Assume that $P(X_i = t_0) > 0$. Then:

$$\begin{aligned} E \left[F_X(X_i) \prod_{j \neq i} F_{X_j}(X_i) \right] &= \int F_X(X_i) \prod_{j \neq i} F_{X_j} dF_{X_i} \\ &= \int_{\mathbb{R} \setminus \{t_0\}} F_X(X_i) \prod_{j \neq i} F_{X_j} dF_{X_i} + \int_{\{t_0\}} F_X(X_i) \prod_{j \neq i} F_{X_j} dF_{X_i} \\ &\leq \int_{\mathbb{R} \setminus \{t_0\}} \prod_{j=1}^n F_{X_j} dF_{X_i} + P(X_i = t_0) F_X(X_i)(t_0) \prod_{j \neq i} F_{X_j}(t_0) \\ &< \int_{\mathbb{R} \setminus \{t_0\}} \prod_{j=1}^n F_{X_j} dF_{X_i} + P(X_i = t_0) \prod_{j=1}^n F_{X_j}(t_0) \\ &= \int_{\mathbb{R} \setminus \{t_0\}} \prod_{j=1}^n F_{X_j} dF_{X_i} + \int_{\{t_0\}} \prod_{j=1}^n F_{X_j} dF_{X_i} \\ &= E \left[\prod_{j=1}^n F_{X_j}(X_i) \right]. \end{aligned}$$

- Assume now that there is not t_0 satisfying the conditions and such that $P(X_i = t_0) = 0$. By the previous lemma, there is an interval $[a, b]$ and $\varepsilon > 0$ such that

$$\prod_{j=1}^n F_{X_j}(t) - F_X(t) \prod_{j \neq i} F_{X_j}(t) \geq \varepsilon > 0$$

for any $t \in [a, b]$ and $P(X_i \in [a, b]) > 0$. Then:

$$\begin{aligned}
E \left[F_X(X_i) \prod_{j \neq i} F_{X_j}(X_i) \right] &= \int F_X(X_i) \prod_{j \neq i} F_{X_j} dF_{X_i} \\
&= \int_{\mathbb{R} \setminus [a, b]} F_X(X_i) \prod_{j \neq i} F_{X_j} dF_{X_i} + \int_{[a, b]} F_X(X_i) \prod_{j \neq i} F_{X_j} dF_{X_i} \\
&\leq \int_{\mathbb{R} \setminus [a, b]} \prod_{j=1}^n F_{X_j} dF_{X_i} + \int_{[a, b]} \left(\prod_{j=1}^n F_{X_j} - \varepsilon \right) dF_{X_i} \\
&= \int \prod_{j=1}^n F_{X_j} dF_{X_i} + \varepsilon P(X_i \in [a, b]) < E \left[\prod_{j=1}^n F_{X_j}(X_i) \right]. \blacksquare
\end{aligned}$$

We have seen that $X \succeq_{\text{FSD}} X_i$ for any $i = 1, \dots, n$, implies that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ when the random variables are independent. Since general statistical preference is based on the joint distribution, and as a consequence takes into account the possible stochastic dependencies between the variables, we are going to study a number of cases where the variables are not independent. In the remainder of this subsection we shall focus on comonotonic random variables.

In Equation (3.6) of Proposition 3.16 we saw that the probabilistic relation $Q(X, Y)$ for two continuous and comonotonic random variables is given by:

$$Q(X, Y) = \int_{x: F_X(x) < F_Y(x)} f_X(x) dx + \frac{1}{2} \int_{x: F_X(x) = F_Y(x)} f_X(x) dx,$$

where f_X denotes the density function of X .

In a similar manner, we can extend this expression to the functional $Q_n(\cdot, [\cdot])$. In order to do this, we must first introduce the notion of Dirac-delta functional. Let us consider the function $H_a : \mathbb{R} \rightarrow [0, 1]$ given by:

$$H_a(x) = \begin{cases} 0 & \text{if } x < a. \\ 1 & \text{if } x \geq a. \end{cases}$$

The Dirac-delta functional δ_a (see [66]) associated to H_a is an application that satisfies:

- $\delta_a(t) = 0$ for every $t \neq a$ and
- $\int_{\mathbb{R}} \delta_a(t) dt = 1$.

In such a case, it holds that:

$$H_a(x) = \int_{-\infty}^x \delta_a(t) d(t) \text{ for every } x \in \mathbb{R}. \quad (3.27)$$

This functional is not a real-valued function because it does not take a real value in a . It plays the role of the density function for a probability distribution that takes the value a with probability 1, and we shall use it in the proof of the following lemma.

Lemma 3.94 *Let X, X_1, \dots, X_n be absolutely continuous and comonotonic real-valued random variables satisfying Equation (3.25). Then*

$$Q_n(X, [X_1, \dots, X_n]) = \int_{x: F_X(x) < F_{X_1}(x), \dots, F_{X_n}(x)} f_X(x) d(x).$$

Proof: By Corollary 3.89, it holds that:

$$Q_n(X, [X_1, \dots, X_n]) = P(X > \max(X_1, \dots, X_n)).$$

Since the random variables are comonotonic, their joint distribution function F is given by:

$$F(x, x_1, \dots, x_n) = \min(F_X(x), F_{X_1}(x_1), \dots, F_{X_n}(x_n))$$

for every $x, x_1, \dots, x_n \in \mathbb{R}$. Let us compute the distribution function of $\max(X_1, \dots, X_n)$ and X , denoted by F^* :

$$\begin{aligned} F^*(x, y) &= P(X \leq x, \max(X_1, \dots, X_n) \leq y) \\ &= P(X \leq x, X_1 \leq y, \dots, X_n \leq y) = F(x, y, \dots, y). \end{aligned}$$

Thus, this distribution function can be expressed by:

$$\begin{aligned} F^*(x, y) &= F(x, y, \dots, y) = \min(F_X(x), F_{X_1}(y), \dots, F_{X_n}(y)) \\ &= \begin{cases} F_X(x) & \text{if } F_X(x) \leq \min(F_{X_1}(y), \dots, F_{X_n}(y)). \\ \min(F_{X_1}(y), \dots, F_{X_n}(y)) & \text{if } F_X(x) > \min(F_{X_1}(y), \dots, F_{X_n}(y)). \end{cases} \end{aligned}$$

Equivalently,

$$F^*(x, y) = \begin{cases} F_X(x) & \text{if } y \geq h^{-1}(F_X(x)), \\ \min(F_{X_1}(y), \dots, F_{X_n}(y)) & \text{if } y < h^{-1}(F_X(x)), \end{cases}$$

where h^{-1} denotes the pseudo-inverse of the function h given by:

$$h(y) = \min(F_{X_1}(y), \dots, F_{X_n}(y)) \text{ for every } y \in \mathbb{R}.$$

Note that the pseudo-inverse is well-defined since h is an increasing function. Now, $\frac{\partial F^*}{\partial x}(x, y) = 0$ for every (x, y) satisfying $y < h^{-1}(F_X(x))$. Moreover, if we restrict to the points (x, y) such that $y \geq h^{-1}(F_X(x))$, we obtain that:

$$\frac{\partial F^*}{\partial x}(x, y) = f_X(x).$$

Thus, if we assume that:

$$\frac{\partial F^*}{\partial x}(x, y) = \begin{cases} 0 & \text{if } y < h^{-1}(F_X(x)), \\ f_X(x) & \text{if } y \geq h^{-1}(F_X(x)), \end{cases}$$

then:

$$\frac{\partial^2 F^*}{\partial x \partial y}(x, y) = f_X(x) \delta(y - h^{-1}(F_X(x))).$$

As this distribution plays the role of the density function of $\max(X_1, \dots, X_n)$ and X , using Equation (3.27) we can compute the value of $Q_n(X, [X_1, \dots, X_n])$:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= P(X > \max(X_1, \dots, X_n)) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_X(x) \delta(y - h^{-1}(F_X(x))) I_{x>y}(y) dy dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_X(x) \delta(y - h^{-1}(F_X(x))) \lim_n I_{\{x-y \geq 1/n\}}(y) dy dx \\ &= \int_{\mathbb{R}} \lim_n \int_{-\infty}^{x-1/n} f_X(x) \delta(y - h^{-1}(F_X(x))) dy dx \\ &= \int_{\mathbb{R}} \lim_n f_X(x) I_{\{x-1/n \geq h^{-1}(F_X(x))\}}(x) dy dx \\ &= \int_{\mathbb{R}} f_X(x) I_{\{x > h^{-1}(F_X(x))\}}(x) dy dx \\ &= \int_{\{F_X(x) < F_{X_1}(x), \dots, F_{X_n}(x)\}} f_X(x) dx, \end{aligned}$$

where the last equality holds applying the Theorem of Monotone Convergence. \blacksquare

Theorem 3.95 *Let X, X_1, \dots, X_n be $n+1$ absolutely continuous and comonotonic real-valued random variables satisfying Equations (3.25) and (3.26). If $X \succeq_{\text{FSD}} X_i$ for $i = 1, \dots, n$, then $X \succ_{\text{SP}} [X_1, \dots, X_n]$. Moreover, in that case $Q_n(X, [X_1, \dots, X_n]) = 1$.*

Proof: Since $X \succeq_{\text{FSD}} X_i$ for every $i = 1, \dots, n$, then $F_X(x) \leq F_{X_i}(x)$ for every $x \in \mathbb{R}$ and $i = 1, \dots, n$. Applying the previous lemma we obtain that:

$$Q_n(X_i, [X, X_{-i}]) = \int_{\{F_{X_i}(x) < F_X(x), F_{X_j}(x): j \neq i\}} f_{X_i}(x) dx = 0.$$

Thus, $Q_n(X, [X_1, \dots, X_n]) > Q_n(X_i, [X, X_{-i}]) = 0$ for every $i = 1, \dots, n$. Since

$$Q_n(X, [X_1, \dots, X_n]) + \sum_{i=1}^n Q_n(X_i, [X, X_{-i}]) = 1,$$

it holds that:

$$Q_n(X, [X_1, \dots, X_n]) = 1.$$

Then, $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. ■

Let us now investigate the case in which the random variables X, X_1, \dots, X_n are comonotonic and discrete with finite supports. When $n = 1$, De Meyer et al. proved (see Proposition 3.20) that the supports of the variables can be expressed by $S_X = \{x_1, \dots, x_m\}$ and $S_{X_1} = \{x_1^{(1)}, \dots, x_m^{(1)}\}$ such that

$$P(X = x_i, X_1 = x_i^{(1)}) = P(X = x_i) = P(X_1 = x_i^{(1)}) \text{ for any } i = 1, \dots, m.$$

We are going to prove the a similar expression can be found when $n \geq 2$.

Lemma 3.96 *Let X, X_1, \dots, X_n be $n + 1$ discrete and countermonotonic real-valued random variables with finite supports. Then, their supports can be expressed by*

$$S_X = \{x_1, \dots, x_m\}, S_{X_1} = \{x_1^{(1)}, \dots, x_m^{(1)}\}, \dots, S_{X_n} = \{x_1^{(n)}, \dots, x_m^{(n)}\}, \quad (3.28)$$

and

$$P(X = x_i, X_1 = x_i^{(1)}, \dots, X_n = x_i^{(n)}) = P(X = x_i) = \dots = P(X_n = x_i^{(n)}), \quad (3.29)$$

for any $i = 1, \dots, n$.

Proof: We apply induction on n . First of all, when $n = 1$, this lemma becomes Proposition 3.20. Assume then that the result holds for $n - 1$. Consider the variables X, X_1, \dots, X_n . Apply the induction hypothesis on X, X_1, \dots, X_{n-1} . Then, the supports of these variables can be expressed as in Equation (3.28), and they also satisfy Equation (3.29). Now, apply Proposition 3.20 to X (with the new support) and X_n . Then, if in this process we duplicate an element x_i , we also duplicate the elements $x_i^{(j)}$ for any $j = 1, \dots, n - 1$, and we adapt the probabilities in order to obtain the equalities:

$$P(X = x_i) = P(X_n = x_i^{(1)}) = \dots = P(X_n = x_i^{(n)}).$$

Finally, let us prove that

$$P(X = x_i, X_1 = x_i^{(1)}, \dots, X_n = x_i^{(n)}) = P(X = x_i).$$

For this aim, note that

$$F_X(x_j) = P(X = x_1) + \dots + P(X = x_j) = P(X_i = x_1^{(i)}) + \dots + P(X_i = x_j^{(i)}) = F_{X_i}(x_j^{(i)})$$

for any $j = 1, \dots, m$ and $i = 1, \dots, n$. Then:

$$\begin{aligned} F_{X, X_1, \dots, X_n}(x_{i_0}, x_{i_1}^{(1)}, \dots, x_{i_n}^{(n)}) &= \min(F_X(x_{i_0}), F_{X_1}(x_{i_1}^{(1)}), \dots, F_{X_n}(x_{i_n}^{(n)})) \\ &= \min(F_X(x_{i_0}), F_X(x_{i_1}^{(1)}), \dots, F_X(x_{i_n}^{(n)})) \\ &= F_X(\min_{k=0, \dots, n}(x_{i_k})). \end{aligned}$$

In particular, when $i_0 = i_1 = \dots, i_n$, the previous expression becomes:

$$F_{X, X_1, \dots, X_n}(x_i, x_i^{(1)}, \dots, x_i^{(n)}) = F_X(x_i).$$

Now, consider $(x_{i_0}, x_{i_1}^{(1)}, \dots, x_{i_n}^{(n)})$, and assume that there are k, l such that $i_k \neq i_l$. Since in the proof of Proposition 3.20 (see [54, Proposition 2]) it is showed that $P(X_k = x_{i_k}^{(k)}, X_l = x_{i_l}^{(l)}) = 0$, we deduce that:

$$P(X = x_{i_0}, X_1 = x_{i_1}^{(1)}, \dots, X_n = x_{i_n}^{(n)}) \leq P(X_k = x_{i_k}^{(k)}, X_l = x_{i_l}^{(l)}) = 0.$$

Consequently:

$$\begin{aligned} P(X = x_i, X_i = x_i^{(1)}, \dots, X_i = x_i^{(n)}) &= F(x_i, x_i^{(1)}, \dots, x_i^{(n)}) - F(x_{i-1}, x_{i-1}^{(1)}, \dots, x_{i-1}^{(n)}). \\ &= F_X(x_i) - F_X(x_{i-1}) = P(X = x_i). \quad \blacksquare \end{aligned}$$

Next result gives an expression of the probabilistic relation, generalizing Equation (3.8).

Proposition 3.97 *Consider $n+1$ discrete and comonotonic real-valued random variables X, X_1, \dots, X_n with finite supports. Applying the previous lemma, we can assume that the supports are expressed as in Equation (3.28) satisfying Equation (3.29). Then:*

$$Q_n(X, [X_1, \dots, X_n]) = \sum_{i=1}^n P(X = x_i) \delta_i,$$

where

$$\delta_i = \begin{cases} 0, & \text{if } x_i > x_i^{(1)}, \dots, x_i^{(n)}. \\ \frac{1}{2}, & \text{if } x_i = x_i^{(j)} > x_i^{(k)}, \text{ for any } k \neq j. \\ \frac{1}{3}, & \text{if } x_i = x_i^{(j_1)} = x_i^{(j_2)} > x_i^{(k)}, \text{ for any } k \neq j_1, j_2. \\ \dots & \\ \frac{1}{n}, & \text{if } x_i = x_i^{(1)} = \dots = x_i^{(n)}. \end{cases}$$

Proof: Taking into account Equation (3.29), it holds that:

$$\begin{aligned} P(X > X_1, \dots, X_n) &= \sum_{i_0=1}^m \dots \sum_{i_n=1}^m P(X = x_{i_0}, X_1 = x_{i_1}^{(1)}, \dots, X_n = x_{i_n}^{(n)}) I_{x_{i_0} > x_{i_1}^{(1)}, \dots, x_{i_n}^{(n)}} \\ &= \sum_{i=1}^m P(X = x_i, X_1 = x_i^{(1)}, \dots, X_n = x_i^{(n)}) I_{x_i > x_i^{(1)}, \dots, x_i^{(n)}}. \end{aligned}$$

Similarly:

$$\begin{aligned} P(X = X_{l_1} = \dots = X_{l_k} > X_j : j \neq l_1, \dots, l_k) \\ = \sum_{i_0=1}^m \dots \sum_{i_n=1}^m P(X = x_{i_0}, X_1 = x_{i_1}^{(1)}, \dots, X_n = x_{i_n}^{(n)}) I_A \\ = \sum_{i=1}^m P(X = x_i, X_1 = x_i^{(1)}, \dots, X_n = x_i^{(n)}) I_B, \end{aligned}$$

where A and B are defined by:

$$\begin{aligned} A &= \{x_{i_0} = x_{i_1}^{(1)} = \dots = x_{i_k}^{(k)} > x_{i_j}^{(j)} : j \neq l_1, \dots, l_k\} \text{ and} \\ B &= \{x_i = x_i^{(1)} = \dots = x_i^{(k)} > x_i^{(j)} : j \neq l_1, \dots, l_k\}. \end{aligned}$$

Then:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= \\ &= \sum_{\substack{k=0, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n}} \frac{1}{k+1} P(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i_1, \dots, i_k} (X_j)) \\ &= \sum_{i=1}^m \sum_{i=1}^n P(X = x_i) \delta_i. \quad \blacksquare \end{aligned}$$

Remark 3.98 *In this result we have not imposed Equations (3.25) and (3.26), and thus, it is applicable for all discrete comonotonic random variables with finite supports. ♦*

Using this lemma, we can prove that when the random variables are comonotonic and discrete with finite supports, first degree stochastic dominance also implies general statistical preference.

Theorem 3.99 *Let X, X_1, \dots, X_n be $n + 1$ discrete comonotonic real-valued random variables with finite supports. Then $X \succeq_{\text{FSD}} X_i$ for $i = 1, \dots, n$ implies $X \succeq_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: Using the previous lemma, the supports of X, X_1, \dots, X_n can be expressed as in Equation (3.28) satisfying Equation (3.29). If $X \succeq_{\text{FSD}} X_i$, we have seen in the proof of Theorem 3.51 that $\{i : x_i < x_i^{(j)}\} = \emptyset$ for $j = 1, \dots, n$. Using the previous proposition:

$$\begin{aligned} Q_n(X_i, [X, X_{-i}]) &= \\ &= \sum_{\substack{k=0, \dots, n-1 \\ 1 \leq i_1 < \dots < i_k \leq n}} \frac{1}{k+1} P(X_i = X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i, i_1, \dots, i_{k-1}} (X_j)) \\ &\leq \sum_{\substack{k=0, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n}} \frac{1}{k+1} P(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i_1, \dots, i_k} (X_j)) = Q(X, Y), \end{aligned}$$

and this for any $i = 1, \dots, n$. Then, $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. \blacksquare

Finally, let us prove that when X is strictly preferred to any X_i with respect to first degree stochastic dominance, it is also preferred to $[X_1, \dots, X_n]$ with respect to the general statistical preference.

Proposition 3.100 *Let X, X_1, \dots, X_n be $n + 1$ discrete comonotonic real-valued random variables with finite supports. Then $X \succ_{\text{FSD}} X_i$ for $i = 1, \dots, n$ implies $X \succ_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: Let us prove that $Q_n(X, [X_1, \dots, X_n]) > Q_n(X_i, [X, X_{-i}])$ for $i = 1, \dots, n$. From the proof of the previous result, it suffices to prove that there are k and l such that

$$x_k = x_k^{(j_1)} = \dots = x_k^{(j_l)} > x_k^{(i)}, x_k^{(j)}, \text{ such that } j \neq i, j_1, \dots, j_l.$$

Since $X \succ_{\text{FSD}} X_i$, there is $x_k^{(i)}$ such that $F_X(x_k^{(i)}) < F_{X_i}(x_k^{(i)})$. Furthermore:

$$F_{X_i}(x_k^{(i)}) = P(X_i = x_1^{(i)}) + \dots + P(X_i = x_k^{(i)}) = P(X = x_1) + \dots + P(X = x_k) = F_X(x_k).$$

Then, $x_k > x_k^{(i)}$. Then, there is l such that

$$x_k = x_k^{(j_1)} = \dots = x_k^{(j_l)} > x_k^{(i)}, x_k^{(j)}, \text{ such that } j \neq i, j_1, \dots, j_l,$$

and this proves that $Q_n(X, [X_1, \dots, X_n]) > Q_n(X_i, [X, X_{-i}])$, for $i = 1, \dots, n$. Hence $X \succ_{\text{SP}} [X_1, \dots, X_n]$. \blacksquare

Random variables coupled by Archimedean copulas

Consider $n + 1$ absolutely continuous random variables X, X_1, \dots, X_n coupled by an Archimedean copula C with generator φ . In that case, Equation (2.9) implies that the joint distribution function, F , is given by:

$$F(x, x_1, \dots, x_n) = \varphi^{-1}(\varphi(F_X(x)) + \varphi(F_{X_1}(x_1)) + \dots + \varphi(F_{X_n}(x_n))).$$

Let us try to differentiate this function.

$$\begin{aligned} \frac{\partial F}{\partial x}(x, x_1, \dots, x_n) &= \\ &= \left(\varphi^{-1}\right)' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x_1)) + \dots + \varphi(F_{X_n}(x_n))\right) \varphi'(F_X(x)) f_X(x). \end{aligned}$$

Note that $(\varphi^{-1})'(t)$ equals $(\varphi^{-1})'(t)$ whenever $t \in [0, \varphi(0))$, and $(\varphi^{-1})'(t) = 0$ otherwise. If we continue differentiating with respect to x_1, \dots, x_n , we obtain the following

expression:

$$\begin{aligned} \frac{\partial^2 F}{\partial x \partial x_1}(x, x_1, \dots, x_n) &= (\varphi^{-1})'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x_1)) + \dots + \varphi(F_{X_n}(x_n)) \right) \\ &\quad \cdot \varphi'(F_X(x)) \varphi'(F_{X_1}(x_1)) f_X(x) f_{X_1}(x_1). \\ &\quad \dots \\ \frac{\partial^{n+1} F}{\partial x \partial x_1 \dots \partial x_n}(x, x_1, \dots, x_n) &= (\varphi^{-1})^{(n+1)} \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x_1)) + \dots \right. \\ &\quad \left. + \varphi(F_{X_n}(x_n)) \right) \varphi'(F_X(x)) \prod_{i=1}^n \varphi'(F_{X_i}(x_i)) f_{X_i}(x_i) f_X(x). \end{aligned}$$

Thus, function $f(x, x_1, \dots, x_n) = \frac{\partial^{n+1} F}{\partial x \partial x_1 \dots \partial x_n}(x, x_1, \dots, x_n)$ is the density function of X, X_1, \dots, X_n whenever $f \neq 0$, since it is the $n+1$ derivative of F , and the $n+1$ integral over \mathbb{R}^{n+1} equals 1. In addition, f becomes the density function of Equation (3.10). Note that $f \neq 0$ when $(\varphi^{-1})^{(n+1)}(t) > 0$ for some $t \in \mathbb{R}$. Moreover, if f is the joint density, $P(X = X_i) = P(X_i = X_j) = 0$ for every i, j ($i \neq j$). Consequently, for such variables it holds that:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= P(X > \max(X_1, \dots, X_n)) \\ &= P(X \geq \max(X_1, \dots, X_n)) = Q(X, \max(X_1, \dots, X_n)). \end{aligned}$$

Using the joint density function f , we can prove the following result.

Theorem 3.101 *Let X, X_1, \dots, X_n be $n+1$ absolutely continuous random variables coupled by an Archimedean copula C generated by φ , that satisfies $(\varphi^{-1})^{(n+1)} \neq 0$. Then, if $X \succeq_{\text{FSD}} X_i$ for every $i = 1, \dots, n$, then $X \succeq_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: We know that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ if and only if

$$P(X \geq \max(X_1, \dots, X_n)) \geq P(X_i \geq \max(X, X_{-i})),$$

for every $i = 1, \dots, n$. Let us compute $P(X \geq \max(X_1, \dots, X_n))$.

$$\begin{aligned} P(X \geq \max(X_1, \dots, X_n)) &= \int_{\mathbb{R}} \int_{-\infty}^x \dots \int_{-\infty}^x f(x, x_1, \dots, x_n) dx_n \dots dx_1 dx \\ &= \int_{\mathbb{R}} (\varphi^{-1})' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi'(F_X(x)) f_X(x) dx. \end{aligned}$$

If we consider

$$\begin{aligned} u &= (\varphi^{-1})' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right), \\ dv &= \varphi'(F_X(x)) f_X(x) dx, \end{aligned}$$

and we make a change of variable, we obtain the following expression:

$$\begin{aligned} P(X \geq \max(X_1, \dots, X_n)) &= \\ &= 1 - \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi'(F_X(x)) \\ &\quad \cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx. \end{aligned}$$

Now, since $X \succeq_{\text{FSD}} X_i$, then $F_X \leq F_{X_i}$, and consequently, as $\varphi(F_X(x)) \geq \varphi(F_{X_i}(x))$ (φ is decreasing), φ' is negative and $(\varphi^{-1})''$ is positive, it holds that:

$$\begin{aligned} P(X \geq \max(X_1, \dots, X_n)) &\geq \\ &1 - \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_{X_i}(x)) \\ &\cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx. \end{aligned}$$

Following the same lines we can also find the expression of $P(X_i \geq \max(X, X_{-i}))$:

$$\begin{aligned} P(X_i \geq \max(X, X_{-i})) &= \\ &1 - \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_{X_i}(x)) \\ &\cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx. \end{aligned}$$

We conclude that:

$$P(X \geq \max(X_1, \dots, X_n)) \geq P(X_i \geq \max(X, X_{-i})),$$

and consequently $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. ■

Finally, let us see that when the Archimedean copula is strict, strict statistical first degree stochastic dominance also implies strict statistical preference.

Proposition 3.102 *Let X, X_1, \dots, X_n be $n+1$ absolutely continuous random variables coupled by an strict Archimedean copula C generated by φ , that satisfies $(\varphi^{-1})^{(n+1)} \neq 0$. Then, if $X \succ_{\text{FSD}} X_i$ for every $i = 1, \dots, n$, then $X \succ_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: By Lemma 3.48, since $X \succ_{\text{FSD}} X_i$, there is an interval $[a, b]$ such that $F_X(t) < F_{X_i}(t)$ for any $t \in [a, b]$ and $P(X_i \in [a, b]) > 0$. Furthermore, we can assume that F_{X_i} is strictly increasing in such interval (otherwise it suffices to consider a subinterval of $[a, b]$ where this function is strictly increasing).

We have seen in the previous proof that

$$\begin{aligned} P(X \geq \max(X_1, \dots, X_n)) &= \\ &1 - \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_X(x)) \\ &\cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx \end{aligned}$$

and

$$\begin{aligned} P(X_i \geq \max(X, X_{-i})) = \\ 1 - \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_{X_i}(x)) \\ \cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx. \end{aligned}$$

Then, in order to prove that $Q_n(X, [X_1, \dots, X_n]) > Q_n(X_i, [X, X_{-i}])$, it suffices to prove that:

$$\begin{aligned} 1 - \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_X(x)) \\ \cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx \\ > 1 - \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_{X_i}(x)) \\ \cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx, \end{aligned}$$

or equivalently:

$$\begin{aligned} \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_X(x)) \\ \cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx \\ < \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \varphi(F_{X_1}(x)) + \dots + \varphi(F_{X_n}(x)) \right) \varphi(F_{X_i}(x)) \\ \cdot \left(\varphi'(F_X(x)) f_X(x) + \sum_{i=1}^n \varphi'(F_{X_i}(x)) f_{X_i}(x) \right) dx. \end{aligned}$$

By the proof of the previous theorem, we know that:

$$\begin{aligned} \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_X(x))' \varphi(F_X(x)) f_X(x) dx \leq \\ \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_X(x))' \varphi(F_{X_i}(x)) f_{X_i}(x) dx \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_j}(x))' \varphi(F_X(x)) f_{X_j}(x) dx \leq \\ \int_{\mathbb{R}} \left(\varphi^{-1} \right)'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_j}(x))' \varphi(F_{X_i}(x)) f_{X_j}(x) dx. \end{aligned}$$

Now, let us see that for $j = i$, the previous inequality is strict. For any $t \in [a, b]$:

$$\begin{aligned} F_{X_i}(t) < F_X(t) &\stackrel{\varphi \text{ decr.}}{\implies} \varphi(F_{X_i}(t)) > \varphi(F_X(t)) \\ &\stackrel{\varphi' \leq 0}{\implies} \varphi'(F_{X_i}(t))\varphi(F_{X_i}(t)) < \varphi'(F_X(t))\varphi(F_X(t)) \\ &\stackrel{(\varphi^{-1})'' < 0}{\implies} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi'(F_{X_i}(t))\varphi(F_{X_i}(t)) > \\ &\quad (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi'(F_X(t))\varphi(F_X(t)). \end{aligned}$$

Then, there is $\varepsilon > 0$ and $[a_1, b_1] \subseteq [a, b]$ such that

$$\begin{aligned} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi'(F_{X_i}(t))\varphi(F_{X_i}(t)) - \\ (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi'(F_X(t))\varphi(F_X(t)) \geq \varepsilon > 0 \end{aligned}$$

for any $t \in [a_1, b_1]$. Then:

$$\begin{aligned} \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_i}(x))' \varphi(F_{X_i}(x)) f_{X_i}(x) dx = \\ \int_{\mathbb{R} \setminus [a_1, b_1]} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_i}(x))' \varphi(F_{X_i}(x)) f_{X_i}(x) dx \\ + \int_{[a_1, b_1]} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_i}(x))' \varphi(F_{X_i}(x)) f_{X_i}(x) dx \\ \geq \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_i}(x))' \varphi(F_X(x)) f_{X_i}(x) dx + \\ \int_{[a_1, b_1]} \varepsilon f_{X_i}(x) dx = \\ \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_i}(x))' \varphi(F_X(x)) f_{X_i}(x) dx + \\ \varepsilon P(X_i \in [a_1, b_1]) > \\ \int_{\mathbb{R}} (\varphi^{-1})'' \left(\varphi(F_X(x)) + \sum_{k=1}^n \varphi(F_{X_k}(x)) \right) \varphi(F_{X_i}(x))' \varphi(F_X(x)) f_{X_i}(x) dx. \end{aligned}$$

Therefore, $Q_n(X, [X_1, \dots, X_n]) > Q_n(X_i, [X, X_{-i}])$, and then we can conclude that $X \succ_{\text{SP}} [X_1, \dots, X_n]$. \blacksquare

We have seen several situations where $X \succeq_{\text{FSD}} X_i \forall i = 1, \dots, n$ implies $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. However, this implication does not hold in general, as we can see in the following example.

Example 3.103 We have seen in Example 3.43 two random variables X and Y such that $X \succ_{\text{FSD}} Y$ and $Y \succ_{\text{SP}} X$. These random variables were defined by:

X/Y	0	1	2
0	0.2	0.15	0
1	0	0.2	0.15
2	0.2	0	0.1

It holds that $Q(X, Y) = 0.45$. Let us modify this example to show that if there is a random variable X that stochastically dominates any other random variables, it may not be the preferred with respect to the general statistical preference. Consider X_1, \dots, X_n equally distributed such that they take a fixed value $c < 0$ with probability 1. Since X and Y greater than X_1, \dots, X_n with probability one, $X \succ_{\text{FSD}} X_i$ for $i = 1, \dots, n$, and it holds that:

$$\begin{aligned} Q_{n+1}(X, [Y, X_1, \dots, X_n]) &= P(X > \max(Y, X_1, \dots, X_n)) \\ &\quad + \frac{1}{2}P(X = Y > \max(X_1, \dots, X_n)) \\ &= P(X > Y) + \frac{1}{2}P(X = Y) = Q(X, Y) = 0.45. \end{aligned}$$

Similarly, $Q_{n+1}(Y, [X, X_1, \dots, X_n]) = Q(Y, X) = 0.55$. Therefore, $X \succ_{\text{FSD}} Y$, $X \succ_{\text{FSD}} X_i$ for $i = 1, \dots, n$ but $X \not\succeq_{\text{SP}} [Y, X_1, \dots, X_n]$. ♦

To conclude this section we are going to see that if we relax the conditions of Theorems 3.91, 3.95, 3.99 or 3.101, then statistical preference does not hold in general. In particular, we replace the hypothesis $X \succeq_{\text{FSD}} X_i$ by $X \succeq_{\text{SP}} X_i$ for some i , and we prove that X is not necessarily the preferred variable.

Example 3.104 Consider the absolutely continuous random variables X, X_1, \dots, X_n , whose density functions are given by:

$$\begin{aligned} f_X(t) &= I_{(2,3)} \\ f_{X_1}(t) &= 0.6 \cdot I_{(1,2)}(t) + 0.4 \cdot I_{(3,4)}(t). \\ f_{X_2}(t) &= I_{(2,3)} \\ f_{X_i}(t) &= I_{(0,1)} \text{ for any } i = 3, \dots, n. \end{aligned}$$

It holds that $X \succeq_{\text{SP}} X_i$ for every $i = 1, \dots, n$ and $X \succeq_{\text{FSD}} X_i$ for every $i = 2, \dots, n$, but $X \not\succeq_{\text{FSD}} X_1$. Moreover,

$$\begin{aligned} Q_n(X_1, [X, X_{-1}]) &= P(X_1 \in (3, 4)) = 0.4. \\ Q_n(X, [X_1, \dots, X_n]) &= Q(X_2, [X, X_{-2}]). \\ Q_n(X_i, [X, X_{-i}]) &= 0 \text{ for any } i = 3, \dots, n. \end{aligned}$$

Since the sum of these values is 1:

$$Q_n(X, [X_1, \dots, X_n]) = Q(X_2, [X, X_{-2}]) = \frac{1}{2}(1 - Q_n(X_1, [X, X_{-1}])) = 0.3,$$

and therefore X_1 is not the preferred random variable with respect to the general statistical preference. ♦

Thus, Theorems 3.91, 3.95, 3.99 and 3.101 cannot be extended to any general situations.

3.3.4 General statistical preference Vs n^{th} degree stochastic dominance

In the previous section we established conditions for first degree stochastic dominance to imply general statistical preference. Next we shall investigate the possible relationships between the m^{th} degree stochastic dominance and the general statistical preference.

Consider random variables X, X_1, \dots, X_n and assume that $X \geq_{\text{mSD}} X_i$ ($m \geq 2$) for every $i = 1, \dots, n$. We shall study if under those conditions $X \succeq_{\text{SP}} [X_1, \dots, X_n]$. To see that this is not necessarily the case, consider the absolutely continuous random variables whose density functions are given by:

$$\begin{aligned} f_X(t) &= I_{(5,6)}(t). \\ f_{X_1}(t) &= 0.4 \cdot I_{(0,1)}(t) + 0.6 \cdot I_{(6,7)}(t). \\ f_{X_i}(t) &= I_{(-1,0)}(t) \text{ for every } i = 2, \dots, n. \end{aligned}$$

Then $X \geq_{\text{mSD}} X_i$ for every $i = 1, \dots, n$. In fact, $X \succeq_{\text{FSD}} X_i$ for every $i = 2, \dots, n$. However, X is not statistically preferred to $[X_1, \dots, X_n]$:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= P(X > \max(X_1, \dots, X_n)) = P(X_1 \in (0, 1)) = 0.4. \\ Q_n(X_1, [X, X_j : j \neq 1]) &= P(X_1 > \max(X, X_j : j \neq 1)) = P(X_1 \in (6, 7)) = 0.6. \\ Q_n(X_i, [X, X_j : j \neq i]) &= 0 \text{ for any } i = 2, \dots, n. \end{aligned}$$

Note that due to the definition of the density functions, the values of the relation Q_n are independent of the possible dependence among the random variables. Thus, we conclude that, for $m \geq 2$:

$$X \geq_{\text{mSD}} X_i \text{ for every } i = 1, \dots, n \text{ does not imply } X \succeq_{\text{SP}} [X_1, \dots, X_n].$$

Assume on the other hand that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ and let us investigate whether if $X \geq_{\text{mSD}} X_i$ for some $m \geq 1$. To see that this is not the case, consider the absolutely continuous random variables with density functions

$$\begin{aligned} f_X(t) &= 0.4 \cdot I_{(0,1)}(t) + 0.6 \cdot I_{(2,3)}(t). \\ f_{X_i}(t) &= I_{(1,2)}(t) \text{ for every } i = 1, \dots, n. \end{aligned}$$

$X \succeq_{\text{SP}} [X_1, \dots, X_n]$, because:

$$Q_n(X, [X_1, \dots, X_n]) = P(X > \max(X_1, \dots, X_n)) = P(X \in (2, 3)) = 0.6.$$

However, X does not stochastically dominate X_i by the m^{th} degree for any $m \geq 1$, since $F_X(t) > F_{X_i}(t)$ for every $t \in (0, 1)$, and consequently $G_X^m(t) > G_{X_i}^m(t)$ for every $m \geq 2$ and $t \in (0, 1)$.

We conclude that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$ does not imply that exists $m \geq 1$ such that $X \geq_{\text{mSD}} X_i$ for every $i = 1, \dots, n$. This generalises Remark 3.63, where we saw that there is not a general relationship between the n^{th} degree stochastic dominance and the pairwise statistical preference.

Remark 3.105 *Let us note that if X, X_1, \dots, X_n are $n + 1$ random variables such that $X \succeq_{\text{SP}} \max(X_1, \dots, X_n)$ (respectively, $X \geq_{\text{mSD}} \max(X_1, \dots, X_n)$), then $X \succeq_{\text{SP}} X_i$ (respectively, $X \geq_{\text{mSD}} X_i$) for every $i = 1, \dots, n$. \blacklozenge*

To conclude this section, we present this result:

Proposition 3.106 *Given $n + 1$ real-valued random variables X, X_1, \dots, X_n , $X \succeq_{\text{SP}} \max(X_1, \dots, X_n)$ implies that $X \succeq_{\text{SP}} [X_1, \dots, X_n]$.*

Proof: Since $X \succeq_{\text{SP}} \max(X_1, \dots, X_n)$, it holds that

$$Q(X, \max(X_1, \dots, X_n) \geq Q(\max(X_1, \dots, X_n), X).$$

In particular, by Lemma 2.20, we know that

$$P(X > \max(X_1, \dots, X_n)) \geq P(\max(X_1, \dots, X_n) > X),$$

since:

$$\begin{aligned} P(X > \max(X_1, \dots, X_n)) &\geq P(\max(X_1, \dots, X_n) > X) \\ &= \sum_{\substack{k=1, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n \\ i \neq i_1, \dots, i_k}} P(X_i = X_{i_1} = \dots = X_{i_k} > X, \max_{j \neq i_1, \dots, i_k} (X_j)) \\ &\geq \sum_{\substack{k=1, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n \\ i \neq i_1, \dots, i_k}} \frac{1}{k+1} P(X_i = X_{i_1} = \dots = X_{i_k} > X, \max_{j \neq i_1, \dots, i_k} (X_j)). \end{aligned}$$

Then:

$$\begin{aligned} Q_n(X, [X_1, \dots, X_n]) &= \\ &\sum_{\substack{k=0, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n}} \frac{1}{k+1} P(X = X_{i_1} = \dots = X_{i_k} > \max_{j \neq i_1, \dots, i_k} (X_j)) \geq \\ &\sum_{\substack{k=1, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n \\ i \neq i_1, \dots, i_k}} \frac{1}{k+1} P(X_i = X_{i_1} = \dots = X_{i_k} > X, \max_{j \neq i_1, \dots, i_k} (X_j)) + \\ &\sum_{\substack{k=1, \dots, n \\ 1 \leq i_1 < \dots < i_k \leq n \\ i \neq i_1, \dots, i_k}} \frac{1}{k+1} P(X_i = X_{i_1} = \dots = X_{i_k} > X, \max_{j \neq i_1, \dots, i_k} (X_j)) \\ &= Q_n(X_i, [X, X_{-i}]). \end{aligned} \quad \blacksquare$$

Figure 3.5 summarises some of the results of this section. Missing arrows mean that an implication does not hold in general, arrows with references means that such implication holds in the conditions of such references, and arrow without reference means that such implication always holds.

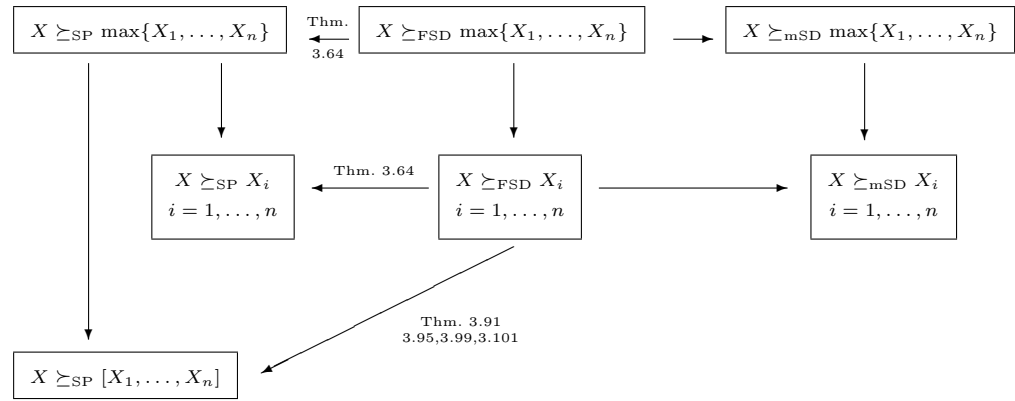


Figure 3.5: Relationships among first and n^{th} degree stochastic dominance, statistical preference and the general statistical preference.

3.4 Applications

In this section we present two possible applications of stochastic orders. On the one hand, we apply stochastic dominance and statistical preference for the comparison of fitness values, and on the other hand, we use the general statistical preference in decision making problems with linguistic variables.

3.4.1 Comparison of fitness values

Genetic algorithms are a powerful tool to perform tasks such as generation of fuzzy rule bases, optimization of fuzzy rule bases, generation of membership functions, and tuning of membership functions (see [41]). All these tasks can be considered as optimization or search processes. A genetic algorithm generates or adapts a fuzzy system, which is called Genetic Fuzzy Systems (GFS, for short) [42]. The use of GFS has been widely accepted,

since these algorithms are robust and can search efficiently large solution spaces (see [213]).

Although in this context the linguistic granules or information are represented by fuzzy sets, the input data and the output results are usually crisp [87]. However, some recent papers (see [180, 181, 182, 183]) have dealt with fuzzy-valued data to learn and evaluate GFS. In that approach the function that quantifies the optimality of a solution in the genetic algorithm, that is, the fitness function, is fuzzy-valued. In particular, in [183], it has been considered that the fitness values are unknown, and that interval valued information is available. The computed fitness value is used by the genetic algorithm module to produce the next population of individuals. In this context some kind of order between two fitness values is necessary if we want to determine whether one individual precedes the other. Since the information about the fitness values is imprecise and is given by means of intervals, a procedure for comparing two intervals is required. Initially, these procedures were based on estimating and comparing two probabilities [183]. In this section we consider statistical preference as a more flexible tool for the comparison of intervals.

Thus, in this section we study of these concepts in connection with the comparison of two intervals, that represent imprecise information about the fitness values of two Knowledge Bases. In particular, we shall make no assumptions about the joint distribution of the two fitness values and shall use then the uniform distribution. This is not an artificial requirement, and it has been considered in many situation as a consequence of lack of information (see, for instance, [183, 197]). When this distribution is considered, we obtain the specific expression of the associated probabilistic and fuzzy relations. We also consider the situation where we have some additional information about the distribution of the fitness, that we model that by means of beta distributions. For these two cases, we consider three possible situations between the intervals: independence, comonotonicity and countermonotonicity.

Usual comparison methods

Let us consider two fitness values θ_1 and θ_2 of two KBs, that is, the mean squared errors of these two KBs on the training set. In many situations, θ_1 and θ_2 are unknown, but we have some imprecise information about them, that we model by means of two intervals that include them. These intervals can be obtained by means of a fuzzy generalisation of the mean squared errors (for a more detailed explanation, see Sections 4 and 5 in [183]) and they will be denoted by $FMSE_1$ and $FMSE_2$, respectively. The comparison of this two intervals is needed in order to choose the predecessor and the successor.

Let us introduce the usual methods that can be found in the literature for the comparison of such intervals. We shall propose statistical preference as an alternative method and investigate the relationships between all the possibilities.

Let us start with the *strong dominance* that was considered in [116]. In that case, if these two intervals are disjoint, then we have not any problem to determine the preferred interval and therefore the decision is trivial. The problem arises when the intersection is non-empty, since the intervals are incomparable.

Definition 3.107 Consider the fitness θ_1 and θ_2 with associated intervals $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$, respectively. It holds that:

- If $b_2 < a_1$, then θ_1 is preferred to θ_2 with respect to the strong dominance, denoted by $\theta_1 \succ_{sd} \theta_2$.
- If $b_1 < a_2$, then θ_2 is preferred to θ_1 with respect to the strong dominance, denoted by $\theta_2 \succ_{sd} \theta_1$.
- Otherwise, θ_1 and θ_2 are incomparable.

This method is too restrictive, since it can be used only in very particular cases. An attempt to solve this problem is to use the first degree stochastic dominance, that introduces prior knowledge about the probability distribution of the fitness.

In particular, if we assume that the fitness follows a uniform distribution (as in [197]), then:

$$\theta_1 \succ_{FSD} \theta_2 \Leftrightarrow a_1 \geq a_2 \text{ and } b_1 \geq b_2,$$

with at least one of the inequalities strict. In particular, if θ_1 strong dominates θ_2 , then $\theta_1 \succ_{FSD} \theta_2$ regardless on the distribution of the fitness.

Nevertheless, first degree stochastic dominance, as we have already noticed during this memory, does not solve all the problems of strong dominance, since, for instance, incomparability is also allowed.

Another method, called *method of the probabilistic prior*, was proposed in [183]. As first degree stochastic dominance, it is based on a prior knowledge about the probability distribution of the fitness, $P(\theta_1, \theta_2)$.

Definition 3.108 Consider the fitness θ_1 and θ_2 with associated intervals $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$. Then, θ_1 is considered to be preferred to θ_2 with respect to the probabilistic prior, and is denoted by $\theta_1 \succ_{pp} \theta_2$, if and only if

$$\frac{P(\theta_1 > \theta_2)}{P(\theta_1 \leq \theta_2)} > 1. \quad (3.30)$$

If $P\{\theta_1 \leq \theta_2\} = 0$, the ration in Equation (3.30) is not defined, but it is assumed that $\theta_1 \succ_{pp} \theta_2$.

Remark 3.109 Recall that from Equation (3.30) we derive that $\theta_1 \succ_{\text{pp}} \theta_2$ if and only if:

$$P(\theta_1 > \theta_2) > P(\theta_1 \leq \theta_2).$$

Thus, the probability prior is equivalent to the probability dominance, with the strict version, considered in Remark 2.22, with $\beta = 0.5$. ♦

Even though these methods allow to compare a wider class of random intervals than the strong dominance, as we said in Remark 2.22 they have an important drawback: they allow for incomparability. In particular, whenever $P(\theta_1 = \theta_2) \geq 0.5$, θ_1 and θ_2 would be incomparable.

Then, it seems natural to consider statistical preference as a method for the comparison of fitness for two main reasons: avoid incomparability and graduate the preference. Also, as we already commented in Subsection 2.1.2, the probabilistic relation Q can be transformed into a fuzzy relation.

Let us study some relationships among strong dominance, first degree stochastic dominance, probabilistic prior and statistical preference.

Proposition 3.110 Given two fitness θ_1 and θ_2 with associated intervals $\text{FMSE}_1 = [a_1, b_1]$ and $\text{FMSE}_2 = [a_2, b_2]$, it holds that:

- $\theta_1 \succ_{\text{sd}} \theta_2$ implies $\theta_1 \succ_{\text{FSD}} \theta_2$.
- $\theta_1 \succ_{\text{sd}} \theta_2$ implies $\theta_1 \succ_{\text{pp}} \theta_2$.
- $\theta_1 \succ_{\text{pp}} \theta_2$ implies $\theta_1 \succ_{\text{SP}} \theta_2$.
- If θ_1 and θ_2 are independent, $\theta_1 \succ_{\text{FSD}} \theta_2$ implies $\theta_1 \succ_{\text{pp}} \theta_2$.

Proof:

- The proof of the first item is based on the fact that $\theta_1 \succ_{\text{sd}} \theta_2$ implies

$$\min \text{FMSE}_1 = a_1 > b_2 \max \text{FMSE}_2,$$

and consequently $\theta_1 \succ_{\text{FSD}} \theta_2$ regardless on the distributions of FMSE_i , $i = 1, 2$.

- If $\theta_1 \succ_{\text{sd}} \theta_2$, then $\{(\theta_1, \theta_2) : \theta_1 \leq \theta_2\} = \emptyset$, and consequently $\theta_1 \succ_{\text{pp}} \theta_2$.
- If $\theta_1 \succ_{\text{pp}} \theta_2$, then $P(\theta_1 > \theta_2) > P(\theta_1 \leq \theta_2)$, that implies $Q(X, Y) > Q(Y, X)$. However, since Q is a probabilistic relation, this means that $Q(X, Y) > \frac{1}{2}$, and thus $\theta_1 \succ_{\text{SP}} \theta_2$.

- If the intervals are independent, then $P(\theta_1 = \theta_2) = 0$, and consequently $\theta_1 \succ_{pp} \theta_2$ if and only if

$$P(\theta_1 > \theta_2) > P(\theta_1 < \theta_2).$$

Thus, both the probabilistic prior and statistical preference are equivalent in this context. Thus, if $\theta_1 \succ_{FSD} \theta_2$, applying Theorem 3.64, $\theta_1 \succ_{SP} \theta_2$, and consequently the preference with respect to the probabilistic prior method also hold. ■

Thus there is a relationship between the probabilistic prior and the stochastic order when the intervals are independent. However, such relationship does not hold for comonotonic and countermonotonic intervals, as we show next:

Example 3.111 Consider θ_1 distributed in the interval $[1, 2]$ and θ_2 distributed in the interval $[0, 2]$. We consider that $FMSE_1$ follows an uniform distribution and the distribution of $FMSE_2$ is defined by the density function:

$$f(x) = \begin{cases} \frac{1}{11} & \text{if } 0 < x < 1.1, \\ 1 & \text{if } 1.1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\theta_1 \succ_{FSD} \theta_2$. Assume that both intervals are comonotonic. Using Equation (3.6) we can compute $P(\theta_1 = \theta_2)$:

$$P(\theta_1 = \theta_2) = \int_{[1.1, 2]} f_X(x) dx = 0.9.$$

Thus, both intervals are incomparable with respect to the probabilistic prior.

Assume now that they are countermonotonic. Using Equation (3.7) we obtain that

$$Q(\theta_1, \theta_2) = F_Y(1.5) = 0.5.$$

Thus, $\theta_1 \not\succeq_{SP} \theta_2$, and consequently, using Proposition 3.110, $\theta_1 \not\succeq_{pp} \theta_2$. ♦

Table 3.6 summarises the general relationships we have seen during this section.

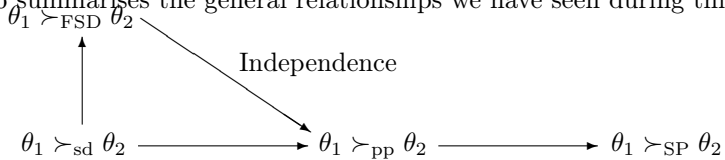


Figure 3.6: Summary of the relationships between strong dominance, first degree stochastic dominance, probabilistic prior and statistical preference given in Proposition 3.110.

Expression of the probabilistic relation for the comparison of fitness values

In this section we will apply statistical preference to the comparison of fitness values.

Uniform case Let us consider again an uniform distribution, that is, no prior information about the distribution over the observed interval, as in [197], and let us search for an expression of the probabilistic relation Q so as to characterise the statistical preference.

Thus, $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$ will denote now two intervals where we know the fitness θ_1 and θ_2 of two KBs are included. Let us assume a uniform distribution on each of them. We will consider again three possible ways to obtain the joint distribution: an assumption of independence, that is, being coupled by the product, and the extreme cases where they are coupled by the minimum or the Łukasiewicz copulas. In these three cases we will obtain the condition on the parameters to assure the statistical preference of the interval $FMSE_1$ to the interval $FMSE_2$. To do that, the expression of the probabilistic relation will be an essential part of the proof.

First of all, recall the result the comparison of independent uniform distributions was already studied in Proposition 3.71: if $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$ be two uniformly distributed intervals which represent the information we have about the fitness θ_1 and θ_2 of two KBs, and the joint distribution is obtained by means of the product copula, then the probabilistic relation $Q(\theta_1, \theta_2)$ takes the following value:

$$Q(\theta_1, \theta_2) = \begin{cases} 1 - \frac{(b_1 - a_2)^2}{2(b_1 - a_1)(b_2 - a_2)} & \text{if } a_1 \leq a_2 < b_1 \leq b_2. \\ 1 - \frac{(b_2 - a_1)^2}{2(b_1 - a_1)(b_2 - a_2)} & \text{if } a_2 \leq a_1 < b_2 \leq b_1. \\ \frac{2b_1 - a_2 - b_2}{2(b_1 - a_1)} & \text{if } a_1 \leq a_2 < b_2 \leq b_1. \\ \frac{b_1 + a_1 - 2a_2}{2(b_2 - a_2)} & \text{if } a_2 \leq a_1 < b_1 \leq b_2. \end{cases}$$

These are the conditions under which $\theta_1 \succ_{SP} \theta_2$:

$$\begin{cases} \text{Always} & \text{if } a_1 \leq a_2 < b_1 \leq b_2. \\ \text{Never} & \text{if } a_2 \leq a_1 < b_2 \leq b_1. \\ a_1 + b_1 \geq b_2 + a_2 & \text{if } a_1 \leq a_2 < b_2 \leq b_1 \\ & \text{or } a_2 \leq a_1 < b_1 \leq b_2. \end{cases}$$

Let us now study the comonotonic case.

Proposition 3.112 *Let $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$ be two uniformly distributed intervals representing the available information on the different fitness θ_1 and θ_2 of two KBs. If the joint distribution is obtained by means of the minimum copula, the*

probabilistic relation $Q(\theta_1, \theta_2)$ takes the following value:

$$Q(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } a_1 \leq a_2 < b_1 \leq b_2. \\ \frac{b_1 - b_2}{b_1 + a_2 - a_1 - b_2} & \text{if } a_1 \leq a_2 < b_2 < b_1. \\ \frac{a_1 - a_2}{b_2 - a_2 - b_1 + a_1} & \text{if } a_2 < a_1 < b_1 \leq b_2. \\ 1 & \text{if } a_2 < a_1 < b_2 \leq b_1. \end{cases}$$

Thus, $\theta_1 \succ_{\text{SP}} \theta_2$ if and only if:

$$\begin{cases} \text{Never} & \text{if } a_1 \leq a_2 < b_1 \leq b_2. \\ \text{Always} & \text{if } a_2 < a_1 < b_2 \leq b_1. \\ a_1 + b_1 > a_2 + b_2 & \text{otherwise.} \end{cases}$$

Then, the condition is equivalent to have a greater expectation.

Proof: The expression of the probabilistic relation can be obtained using Equation (3.6), and taking into account that $P(\theta_1 = \theta_2) = 0$, since the associated cumulative distribution coincide at most in one point.

First and second scenarios of the are trivial. In the third scenario, if $a_1 \leq a_2 < b_2 \leq b_1$ it holds that:

$$\theta_1 \succ_{\text{SP}} \theta_2 \Leftrightarrow \frac{b_1 - b_2}{b_1 + a_2 - a_1 - b_2} > \frac{1}{2} \Leftrightarrow a_1 + b_1 > a_2 + b_2.$$

The condition for $a_2 \leq a_1 < b_1 \leq b_2$ can be similarly obtained. ■

Finally, let us study the countermonotonic case.

Proposition 3.113 *Let $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$ be two uniformly distributed intervals which represent the information we have about the fitness θ_1 and θ_2 of two KBs. If the joint distribution is obtained by means of the Łukasiewicz copula, then the probabilistic relation is given by:*

$$Q(\theta_1, \theta_2) = \frac{b_1 - a_2}{b_2 - a_2 + b_1 - a_1}.$$

In addition, $\theta_1 \succ_{\text{SP}} \theta_2$ if and only if:

$$\begin{cases} \text{Never} & \text{if } a_1 \leq a_2 < b_1 \leq b_2. \\ a_1 + b_1 \geq a_2 + b_2 & \text{if } a_1 \leq a_2 < b_2 < b_1. \\ a_1 + b_1 \geq a_2 + b_2 & \text{if } a_2 < a_1 < b_1 \leq b_2. \\ \text{Always} & \text{if } a_2 < a_1 < b_2 \leq b_1. \end{cases}$$

Proof: The expression of the probabilistic relation can be obtained using Equation (3.7), and taking into account that the point u such that $F_{\theta_1}(u) + F_{\theta_2}(u) = 1$ equals: $u = \frac{b_2 b_1 - a_1 a_2}{b_2 - a_2 + b_1 - a_1}$.

The first and fourth scenarios of the second part are easy, since there they are ordered by means of the stochastic order. In the first scenario it holds that $F_{\theta_1}(u) > F_{\theta_2}(u)$, and consequently

$$Q(\theta_1, \theta_2) < Q(\theta_2, \theta_1),$$

and then $\theta_1 \not\prec_{\text{SP}} \theta_2$. Similarly, we obtain that in the fourth scenario $\theta_1 \succ_{\text{SP}} \theta_2$.

For the second and third scenarios, it is enough to compare the expression of the probabilistic relation with $\frac{1}{2}$. ■

Beta case We now assume that more information about the fitness values may be available. If it is known that some values of the interval are more feasible than others, the uniform distribution is not a good model any more. If we assume that the closer we are to one extreme of the interval the more feasible the values are, beta distributions become more appropriate to model the fitness values. As we made in Subsection 3.2.6, we focus on this situation: beta distributions such that one of the parameters is 1.

As we already said, the density of a beta distribution $\beta(p, q)$ is given by Equation (3.17). However, it is possible to define a beta distribution on every interval $[a, b]$ (it is denoted by $\beta(p, q, a, b)$). The associated density function is:

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(x-a)^{p-1}(b-a)^{q-1}}{(b-a)^{p+q-1}},$$

for any $x \in [a, b]$, and zero otherwise. Next, we will focus on two particular cases. In the first one we will assume that the closer the value is to a_i , the more feasible the value is. In the second case, we will assume the opposite: that the closer the value is to b_i , the more feasible the value is. In terms of density functions, these two cases correspond to strictly decreasing and strictly increasing density functions. We will consider the intervals FMSE_i follows a distribution $\beta(p, 1, a_i, b_i)$, for $i = 1, 2$, where p will be an integer greater than 1. Independently of where the weight of the distribution is, we shall consider three possibilities concerning the relationship between the fitness values: independence, comonotonicity and countermonotonicity. If intervals satisfy one of the following conditions:

$$a_1 \leq a_2 < b_1 \leq b_2 \text{ or } a_2 \leq a_1 < b_2 \leq b_1,$$

we have seen in the previous section that, since they are ordered with respect to the stochastic order, the study of the statistical preference becomes trivial. For this reason we will assume the intervals to satisfy the condition $a_1 \leq a_2 < b_2 \leq b_1$ (the case $a_2 \leq a_1 < b_1 \leq b_2$ can be solved by symmetry).

Proposition 3.114 *Let us consider the different fitness values θ_1 and θ_2 with associated intervals $FMSE_i \equiv [a_i, b_i]$ following a distribution $\beta(p, 1, a_i, b_i)$, where $a_1 \leq a_2 < b_2 \leq b_1$. Then:*

$$Q^P(\theta_1, \theta_2) = p \sum_{k=0}^{p-1} \binom{p-1}{k} \frac{(a_2 - a_1)^{p-k-1} (b_2 - a_2)^{k-1}}{(b_1 - a_1)^p (p+k+1)} + \left(\frac{b_2 - a_1}{b_1 - a_1} \right)^p - \left(\frac{a_2 - a_1}{b_1 - a_1} \right)^p,$$

$$Q^M(\theta_1, \theta_2) = 1 - \left(\frac{t^* - a_1}{b_1 - a_1} \right)^p,$$

$$Q^L(\theta_1, \theta_2) = \left(\frac{z^* - a_2}{b_2 - a_2} \right)^p,$$

where $t^* = \frac{a_1 b_2 - a_2 b_1}{b_2 - a_2 - b_1 + a_1}$ and z^* is the point in $[a_2, b_2]$ such that

$$\left(\frac{z^* - a_1}{b_1 - a_1} \right)^p + \left(\frac{z^* - a_2}{b_2 - a_2} \right)^p = 1,$$

and Q^P , Q^M and Q^L denotes the probabilistic relation when the random variables are coupled by the product, the minimum and the Łukasiewicz operators, respectively.

Proof: Let us begin by computing the expression of $Q^P(\theta_1, \theta_2)$. Since they are independent and continuous, $P(\theta_1 = \theta_2) = 0$. Then:

$$Q^P(\theta_1, \theta_2) = P(\theta_1 > \theta_2) = P(\theta_1 \in [b_2, b_1]) + P(b_2 > \theta_1 > \theta_2).$$

Let us compute each one of the previous probabilities:

$$P(\theta_1 \in [b_2, b_1]) = \int_{a_2}^{b_2} p \frac{(x - a_1)^{p-1}}{(b_1 - a_1)^p} dx = \left(\frac{b_2 - a_1}{b_1 - a_1} \right)^p - \left(\frac{a_2 - a_1}{b_1 - a_1} \right)^p.$$

$$P(b_2 > \theta_1 > \theta_2) = \int_{a_2}^{b_2} \int_{a_2}^x p^2 \frac{(x - a_1)^{p-1}}{(b_1 - a_1)^p} \frac{(y - a_2)^{p-1}}{(b_2 - a_2)^p} dy dx$$

$$= \int_{a_2}^{b_2} p \frac{(x - a_1)^{p-1}}{(b_1 - a_1)^p} \left(\frac{x - a_2}{b_2 - a_2} \right)^p dx.$$

Taking $z = \frac{x - a_2}{b_2 - a_2}$, the previous expression becomes:

$$P(b_2 > \theta_1 > \theta_2) = p \int_0^1 \frac{((b_2 - a_2)z + a_2 - a_1)^{p-1}}{(b_1 - a_1)^p} z^p \frac{dz}{b_2 - a_2}$$

$$= p \int_0^1 \frac{z^p}{(b_2 - a_2)(b_1 - a_1)^p} \sum_{k=0}^{p-1} \binom{p-1}{k} ((b_2 - a_2)z)^k (a_2 - a_1)^{p-1-k} dz$$

$$= \frac{p}{(b_1 - a_1)^p} \sum_{k=0}^{p-1} \binom{p-1}{k} \frac{(a_2 - a_1)^{p-k-1} (b_2 - a_2)^{k-1}}{p+k+1}.$$

Making the sum of the two probabilities, we obtain the value of $Q(\theta_1, \theta_2)$.

Next, assume that θ_1 and θ_2 are comonotonic. Since $\{x : F_{\theta_1}(x) = F_{\theta_2}(x)\} = \emptyset$, applying Equation (3.6) we deduce that

$$Q^M(\theta_1, \theta_2) = \int_{x:F_{\theta_1}(x) < F_{\theta_2}(x)} p \frac{(x - a_1)^{p-1}}{(b_1 - a_1)^p} dx.$$

Moreover, $\{x : F_{\theta_1}(x) < F_{\theta_2}(x)\} = (t^*, b_1]$, where t^* is the point satisfying:

$$\begin{aligned} F_{\theta_1}(t^*) = F_{\theta_2}(t^*) &\Leftrightarrow \left(\frac{t^* - a_1}{b_1 - a_1}\right)^p = \left(\frac{t^* - a_2}{b_2 - a_2}\right)^p \\ &\Leftrightarrow t^*(b_2 - a_2) - a_1(b_2 - a_2) = t^*(b_1 - a_1) - a_2(b_1 - a_1) \\ &\Leftrightarrow t^* = \frac{a_1 b_2 - a_2 b_1}{b_2 - a_2 - b_1 + a_1}. \end{aligned}$$

Then:

$$Q^M(\theta_1, \theta_2) = \int_{t^*}^{b_1} p \frac{(x - a_1)^{p-1}}{(b_1 - a_1)^p} dx = 1 - \left(\frac{t^* - a_1}{b_1 - a_1}\right)^p.$$

Finally, assume that θ_1 and θ_2 are countermonotonic. By Equation (3.7),

$$Q^L(\theta_1, \theta_2) = F_{\theta_2}(z^*) = \left(\frac{z^* - a_2}{b_2 - a_2}\right)^p,$$

where z^* satisfies that:

$$F_{\theta_1}(z^*) + F_{\theta_2}(z^*) = 1 \Leftrightarrow \left(\frac{z^* - a_1}{b_1 - a_1}\right)^p + \left(\frac{z^* - a_2}{b_2 - a_2}\right)^p = 1. \quad \blacksquare$$

Proposition 3.115 *Let us consider the different fitness values θ_1 and θ_2 with associated intervals $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$ following the distribution $\beta(1, q, a_i, b_i)$, where $a_1 \leq a_2 < b_2 \leq b_1$. Then*

$$\begin{aligned} Q^P(\theta_1, \theta_2) &= q \sum_{k=0}^{q-1} \binom{q-1}{k} \frac{(b_1 - b_2)^k (b_2 - a_2)^{q-k-2}}{(b_1 - a_1)^q (q+k+1)} + \left(\frac{b_1 - a_2}{b_1 - a_1}\right)^q, \\ Q^M(\theta_1, \theta_2) &= 1 - \left(\frac{b_1 - t^*}{b_1 - a_1}\right)^q, \\ Q^L(\theta_1, \theta_2) &= 1 - \left(\frac{b_1 - z^*}{b_1 - a_1}\right)^q, \end{aligned}$$

where $t^* = \frac{a_1 b_2 - b_1 a_2}{b_2 - a_2 - b_1 + a_1}$ and z^* is the point in $[a_2, b_2]$ such that

$$\frac{(b_1 - x)^q}{(b_1 - a_1)^{q-1}} + \frac{(b_2 - x)^q}{(b_2 - a_2)^{q-1}} = 1,$$

and Q^P , Q^M and Q^L denotes the probabilistic relation when the random variables are coupled by the product, the minimum and the Łukasiewicz operators, respectively.

Proof: We begin by computing the expression of $Q^P(\theta_1, \theta_2)$. Again, since they are independent and continuous, $P(\theta_1 = \theta_2) = 0$, and then:

$$Q^P(\theta_1, \theta_2) = P(\theta_1 > \theta_2) = P(\theta_1 \in [b_2, b_1]) + P(b_2 > \theta_1 > \theta_2).$$

Let us compute each one the the previous probabilities:

$$\begin{aligned} P(\theta_1 \in [b_2, b_1]) &= \int_{b_2}^{b_1} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} dx = \left(\frac{b_1 - b_2}{b_1 - a_1} \right)^q. \\ P(b_2 > \theta_1 > \theta_2) &= \int_{a_2}^{b_2} \int_{a_2}^x q^2 \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} \frac{(b_2 - y)^{q-1}}{(b_2 - a_2)^q} dy dx \\ &= \int_{a_2}^{b_2} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} \left(1 - \left(\frac{b_2 - x}{b_2 - a_2} \right) \right) dx \\ &= \int_{a_2}^{b_2} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} dx - \int_{a_2}^{b_2} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} \left(\frac{b_2 - x}{b_2 - a_2} \right)^q dx \\ &= \left(\frac{b_1 - a_2}{b_1 - a_1} \right) - \left(\frac{b_1 - b_2}{b_1 - a_1} \right) - \int_{a_2}^{b_2} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} \left(\frac{b_2 - x}{b_2 - a_2} \right)^q dx. \end{aligned}$$

Taking $z = \frac{b_2 - x}{b_2 - a_2}$, the last integral becomes:

$$\begin{aligned} P(b_2 > \theta_1 > \theta_2) &= \int_0^1 q z^q \frac{(b_1 - b_2 + z(b_2 - a_2))^{q-1}}{(b_1 - a_1)^q} \frac{dz}{b_2 - a_2} \\ &= q \int_0^1 \frac{z^q}{(b_2 - a_2)(b_1 - a_1)^q} \sum_{k=0}^{q-1} \binom{q-1}{k} ((b_1 - b_2)z)^k (b_2 - a_2)^{q-k-1} dz \\ &= q \sum_{k=0}^{q-1} \binom{q-1}{k} \frac{(b_1 - b_2)^k (b_2 - a_2)^{q-k-2}}{(b_1 - a_1)^q (q+k+1)} \frac{1}{q+k+1}. \end{aligned}$$

Making the sum of the three terms, we obtain the expression of $Q^P(\theta_1, \theta_2)$.

Consider now the fitness to be comonotonic. Then, since $\{x : F_{\theta_1}(x) = F_{\theta_2}(x)\} = \emptyset$, the expression of the probabilistic relation given in Eq.(3.6) becomes:

$$Q^M(\theta_1, \theta_2) = \int_{x: F_{\theta_1}(x) < F_{\theta_2}(x)} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} dx.$$

Then, $\{x : F_{\theta_1}(x) < F_{\theta_2}(x)\} = (t^*, b_1]$, where:

$$\begin{aligned} F_{\theta_1}(t^*) = F_{\theta_2}(t^*) &\Leftrightarrow 1 - \left(\frac{b_1 - t^*}{b_1 - a_1} \right)^q = 1 - \left(\frac{b_2 - t^*}{b_2 - a_2} \right)^q \\ &\Leftrightarrow \frac{b_1 - t^*}{b_1 - a_1} = \frac{b_2 - t^*}{b_2 - a_2} \Leftrightarrow t^* = \frac{a_1 b_2 - b_1 a_2}{b_2 - a_2 - b_1 + a_1}. \end{aligned}$$

Then:

$$Q^M(\theta_1, \theta_2) = \int_{t^*}^{b_1} q \frac{(b_1 - x)^{q-1}}{(b_1 - a_1)^q} dx = \left(\frac{b_1 - x}{b_1 - a_1} \right)^q.$$

Finally, assume that θ_1 and θ_2 are countermonotonic. Then, $Q^L(\theta_1, \theta_2) = F_{\theta_2}(z^*)$, where z^* satisfies:

$$\begin{aligned} F_{\theta_1}(z^*) + F_{\theta_2}(z^*) = 1 &\Leftrightarrow 1 - \left(\frac{b_1 - x}{b_1 - a_1} \right)^q + 1 - \left(\frac{b_2 - x}{b_2 - a_2} \right)^q = 1 \\ &\Leftrightarrow \left(\frac{b_1 - x}{b_1 - a_1} \right)^q + \left(\frac{b_2 - x}{b_2 - a_2} \right)^q = 1. \quad \blacksquare \end{aligned}$$

Remark 3.116 In order to prove the previous result it is not possible to follow the procedure of Proposition 3.78. There, we used the following property:

$$X \equiv \beta(p, 1) \Leftrightarrow 1 - X \equiv \beta(1, p).$$

Then, since $Q(X, Y) = Q(1 - Y, 1 - X)$ (see Proposition 3.3), the case of $q = 1$ was solved using the case $p = 1$. In the case of general beta distributions, it holds that:

$$X \equiv \beta(p, 1, a, b) \Leftrightarrow (b - a) - X \equiv \beta(1, p, a, b).$$

The problem is that $Q(X, Y) \neq Q((b_2 - a_2) - Y, (b_1 - a_1) - X)$, and therefore this kind of procedure is not possible. \blacklozenge

Remark 3.117 Note that for beta distribution it is not possible to obtain a simpler characterization of the statistical preference like the one for uniform distributions. \blacklozenge

To conclude this section, let us present an example where we show how the values of the probabilistic relation change when we vary the value of p .

Example 3.118 Consider the fitness values θ_1 and θ_2 with associated values $FMSE_1 = [a_1, b_1]$ and $FMSE_2 = [a_2, b_2]$, where $a_1 \leq a_2 < b_2 \leq b_1$, and let assume they follow the beta distribution $\beta(p, 1, a_i, b_i)$. Consider $a_1 = 1, b_1 = 4, a_2 = 2$ and $b_2 = 3$. Table 3.5 shows the values of the probabilistic relation when p moves from 1 to 5, where it is possible to see that θ_1 and θ_2 are equivalent when $p = 1$, but θ_1 is preferred to θ_2 when $p \geq 2$. Moreover, the greater the value of p , the stronger the preference of θ_1 over θ_2 .

Consider now different values of the intervals: $a_1 = 0.7, b_1 = 1.4, a_2 = 0.8$ and $b_2 = 1.2$. In this case, although $[a_2, b_2] \subseteq [a_1, b_1]$ as in the previous example, the difference between b_1 and b_2 is greater than a_1 and a_2 . The results are summarised in Table 3.6. There, we can see that in the three cases, $\theta_1 \succ_{SP} \theta_2$ for any $p \geq 1$. Furthermore, the greater the value of p , the stronger the preference of θ_1 over θ_2 . In Figure 3.7 we can see how the values of Q vary we change the value of the parameter p from 1 to 10. \blacklozenge

p	Q^P	Q^M	Q^L
1	0.5	0.5	0.5
2	0.6853	0.75	0.64
3	0.7945	0.875	0.7436
4	0.8644	0.9375	0.8208
5	0.9101	0.9688	0.8766

Table 3.5: Degrees of preference for the different values of the parameter p for $\text{FMSE}_1 = [1, 4]$ and $\text{FMSE}_2 = [2, 3]$.

p	Q^P	Q^M	Q^L
1	0.5715	0.6667	0.5455
2	0.7076	0.8889	0.64
3	0.7936	0.9630	0.7192
4	0.8533	0.9877	0.7852
5	0.8955	0.9959	0.8384

Table 3.6: Degrees of preference for the different values of the parameter p for $\text{FMSE}_1 = [0.7, 1.4]$ and $\text{FMSE}_2 = [0.8, 1.2]$.

3.4.2 General statistical preference as a tool for linguistic decision making

As we have seen, general statistical preference was introduced as a method that allows for the comparison of more than two random variables. As an illustration of the utility of this method we can consider a decision making problem with linguistic utilities. We consider the example of product management given in [123, Section 8]: a company seeks to plan its production strategy for the next year, and they consider six possible alternatives:

- A_1 : Create a new product for very high-income customers.
- A_2 : Create a new product for high-income customers.
- A_3 : Create a new product for medium-income customers.
- A_4 : Create a new product for low-income customers.
- A_5 : Create a new product suitable for all customers.
- A_6 : Do not create a new product.

Due to the large uncertainty, the three experts of the company are not able to draw the information about the impact of each alternative in a numerical way, and for this reason

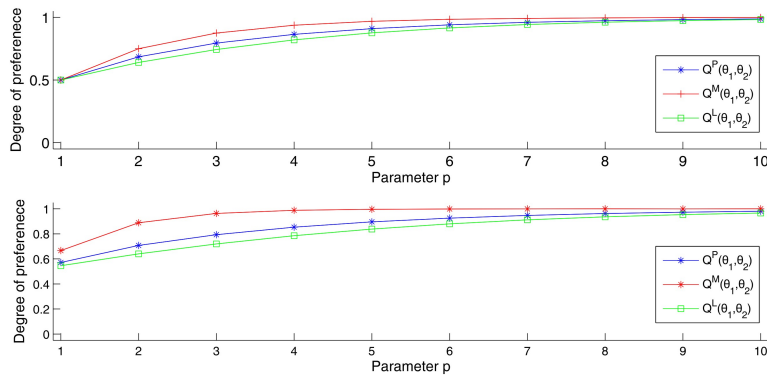


Figure 3.7: Values of the probabilistic relation for different values of p . The above picture corresponds to intervals $[a_1, b_1] = [1, 4]$ and $[a_2, b_2] = [2, 3]$, and the picture below corresponds to intervals $[a_1, b_1] = [0.7, 1.4]$ and $[a_2, b_2] = [0.8, 1.2]$.

they express the utility based on a seven linguistic scale $S = \{s_1, \dots, s_7\}$, where:

- | | |
|------------------|-------------------|
| s_1 : None | s_5 : High |
| s_2 : Very low | s_6 : Very high |
| s_3 : Low | s_7 : Perfect |
| s_4 : Medium | |

Note that the three experts have not the same influence in the company, and its importance is given by the weight vector $(0.2, 0.4, 0.4)$. Moreover, since the decision of each expert depends on the economic situation of the following year, six scenarios are considered:

- | | |
|---------------------|----------------------|
| N_1 : Very bad | N_4 : Regular-Good |
| N_2 : Bad | N_5 : Good |
| N_3 : Regular-Bad | N_6 : Very good |

The experts assume the following weighting vector for these scenarios:

$$W = (0.1, 0.1, 0.1, 0.2, 0.2, 0.3).$$

Finally, the preferences of each expert are given in Tables 3.7, 3.8 and 3.9.

Although in [123] this problem was solved by means of a particular type of aggregation operators, we propose to use the general statistical preference. For any expert e_i , $i = 1, 2, 3$, we can compute the preference degree of the alternative A_j over the others

	N_1	N_2	N_3	N_4	N_5	N_6
A_1	s_2	s_1	s_4	s_6	s_7	s_5
A_2	s_1	s_3	s_5	s_5	s_6	s_6
A_3	s_3	s_4	s_4	s_4	s_4	s_7
A_4	s_2	s_5	s_6	s_4	s_2	s_5
A_5	s_1	s_3	s_4	s_5	s_6	s_6
A_6	s_6	s_5	s_5	s_4	s_2	s_2

Table 3.7: Linguistic payoff matrix-Expert 1.

	N_1	N_2	N_3	N_4	N_5	N_6
A_1	s_3	s_1	s_3	s_5	s_6	s_6
A_2	s_1	s_3	s_4	s_5	s_6	s_6
A_3	s_3	s_4	s_5	s_4	s_3	s_7
A_4	s_3	s_4	s_5	s_4	s_2	s_4
A_5	s_2	s_3	s_4	s_6	s_6	s_6
A_6	s_7	s_6	s_4	s_3	s_2	s_2

Table 3.8: Linguistic payoff matrix-Expert 2.

	N_1	N_2	N_3	N_4	N_5	N_6
A_1	s_1	s_2	s_3	s_5	s_7	s_6
A_2	s_2	s_3	s_4	s_4	s_5	s_6
A_3	s_3	s_4	s_6	s_4	s_3	s_7
A_4	s_2	s_4	s_6	s_4	s_2	s_4
A_5	s_1	s_3	s_4	s_5	s_6	s_6
A_6	s_6	s_6	s_5	s_3	s_2	s_3

Table 3.9: Linguistic payoff matrix-Expert 3.

A_{-j} , and we obtain the following values:

$$\begin{aligned}
Q(A_1, [A_{-1}] | e_1) &= P(N_4) + P(N_5) = 0.4. \\
Q(A_2, [A_{-2}] | e_1) &= 0. \\
Q(A_3, [A_{-3}] | e_1) &= P(N_6) = 0.3. \\
Q(A_4, [A_{-4}] | e_1) &= \frac{1}{2}P(N_2) + P(N_3) = 0.15. \\
Q(A_5, [A_{-5}] | e_1) &= 0. \\
Q(A_6, [A_{-6}] | e_1) &= P(N_1) + \frac{1}{2}P(N_2) = 0.15. \\
Q(A_1, [A_{-1}] | e_2) &= \frac{1}{3}P(N_5) = 0.0667. \\
Q(A_2, [A_{-2}] | e_2) &= \frac{1}{3}P(N_5) = 0.0667. \\
Q(A_3, [A_{-3}] | e_2) &= \frac{1}{2}P(N_3) + P(N_6) = 0.35. \\
Q(A_4, [A_{-4}] | e_2) &= \frac{1}{2}P(N_3) = 0.05. \\
Q(A_5, [A_{-5}] | e_2) &= P(N_4) + \frac{1}{3}P(N_5) = 0.2667. \\
Q(A_6, [A_{-6}] | e_2) &= P(N_1) + P(N_2) = 0.2. \\
Q(A_1, [A_{-1}] | e_3) &= \frac{1}{2}P(N_4) + P(N_5) = 0.3. \\
Q(A_2, [A_{-2}] | e_3) &= 0. \\
Q(A_3, [A_{-3}] | e_3) &= \frac{1}{2}P(N_3) + P(N_6) = 0.35. \\
Q(A_4, [A_{-4}] | e_3) &= \frac{1}{2}P(N_3) = 0.05. \\
Q(A_5, [A_{-5}] | e_3) &= \frac{1}{2}P(N_4) = 0.1. \\
Q(A_6, [A_{-6}] | e_3) &= P(N_1) + P(N_2) = 0.2.
\end{aligned}$$

Now, since the importance of each expert is given by the weighting vector $(0.2, 0.4, 0.4)$, we can obtain the preference degree of each alternative:

$$\begin{aligned}
Q(A_1, [A_{-1}]) &= Q(A_1, [A_{-1}] | e_1)0.2 + Q(A_1, [A_{-1}] | e_2)0.4 \\
&\quad + Q(A_1, [A_{-1}] | e_3)0.4 = 0.4 \cdot 0.2 + 0.0667 \cdot 0.4 + 0.3 \cdot 0.4 = 0.22667.
\end{aligned}$$

And similarly:

$$\begin{aligned}
Q(A_2, [A_{-2}]) &= 0.0667 \cdot 0.4 = 0.02667. \\
Q(A_3, [A_{-3}]) &= 0.3 \cdot 0.2 + 0.35 \cdot 0.4 + 0.35 \cdot 0.4 = 0.34. \\
Q(A_4, [A_{-4}]) &= 0.15 \cdot 0.2 + 0.05 \cdot 0.4 + 0.05 \cdot 0.4 = 0.07. \\
Q(A_5, [A_{-5}]) &= 0.2667 \cdot 0.4 + 0.1 \cdot 0.4 = 0.14667. \\
Q(A_6, [A_{-6}]) &= 0.15 \cdot 0.2 + 0.2 \cdot 0.4 + 0.2 \cdot 0.4 = 0.19.
\end{aligned}$$

Thus, general statistical preference gives A_3 as the preferred alternative: $A_3 \succeq_{SP} [A_{-3}]$; A_1 is the second preferred alternative, A_6 the third, A_5 the fourth, A_4 the fifth and finally A_2 is the less preferred alternative. Consequently, creating a new product for medium-income customers seems to be the best option, while the worst alternative is creating a new product for high-income customers.

3.5 Conclusions

Stochastic orders are tools that allow us to compare random quantities, so they become particularly useful in decision problems under uncertainty. One of the most important stochastic orders that can be found in the literature is stochastic dominance. This method, based on the comparison of the cumulative distribution functions, has been widely studied in the literature, and it has been applied in many different areas. One alternative stochastic order is statistical preference, which has remained unexplored for a long time. For this reason, we have dedicated the first part of this chapter to the investigation of the properties of statistical preference as a stochastic order. In particular, while stochastic dominance is close to the expectation, we have seen that statistical preference is related to another location parameter: the median. This showed that both stochastic orders have a different philosophy under their definition.

Interestingly, there are situations where both stochastic orders give rise to the same conclusions. For instance, we have found conditions under which first degree stochastic dominance implies statistical preference. These situations included, for example, independent random variables or continuous comonotonic/countermonotonic random variables, among others. Although the two methods are not equivalent in general, we have proved that they coincide when comparing independent random variables whose distributions are Bernoulli, exponential, uniform, Pareto, beta and normal.

Both methods have been devised for the pairwise comparison of random variables, and may be unsuitable when more than two random variables must be compared simultaneously. For this reason, we have introduced a new stochastic order, that generalises statistical preference and preserves its underlying philosophy, that allows us to compare more than two random variables at the same time. We have also investigated its main properties and its connection with the usual stochastic orders.

Stochastic orders appear in many different real-life problems. For this reason, the last part of this chapter was devoted to present a number of applications that show the relevance of our results. On the one hand, we have seen that both stochastic dominance and statistical preference could be an interesting alternative to the comparison of fitness values, and on the other hand we have applied the general statistical preference to a multicriteria decision making with linguistic labels.

From the results we have showed in this chapter new open problems arise. For instance, we have given some conditions under which first degree stochastic dominance implies statistical preference, and we have seen that this relation does not hold in general. Thus, a natural question arises: is it possible to characterise the situations in which first degree stochastic dominance implies statistical preference?

Moreover, we have also seen that both stochastic dominance and statistical preference coincide for the comparison of independent random variables whose distribution is

Bernoulli, exponential, normal, ... In fact, both methods reduce to the comparison of the expectation of the variables. We conjecture that for independent random variables whose distribution belongs to the exponential family of distributions, both stochastic dominance and statistical preference coincide and are equivalent to the comparison of the expectation. Although this is an open question that has not been answered yet, a first approach, based on simulations, has already been done by Casero ([32]). We have introduced the general statistical preference as a stochastic order for the comparison of more than two random variables simultaneously. Although we have investigated its main properties, a different approach could be given to this notion. In fact, the general statistical preference could be seen as a fuzzy choice function ([81]) on a set of random variables, since it gives degrees of preference of a random variable over a set of random variables. Then, the investigation of the properties of the general statistical preference as a fuzzy choice function could be an interesting line of research.

4 Comparison of alternatives under uncertainty and imprecision

In the previous chapter we have dealt with the comparison of alternatives under uncertainty. When these alternatives are modelled by means of random variables, the comparison must be performed using stochastic orders. However, there are situations in which it is not possible or adequate to model the experiments by means of a single random variable, due to the presence of imprecision in the experiment. In other words, we focus now in situations where the alternatives are defined under uncertainty but also under imprecision. In such cases, we shall compare sets of random variables instead of single ones; more generally, we shall compare imprecise probability models. For this reason, this chapter is devoted to the extension of the pairwise methods studied in the previous chapter to the comparison of imprecise probability models.

As we have already mentioned, imprecise probabilities ([205]) is a generic term that refers to all mathematical models that serve as an alternative and a generalisation to probability models in case of imprecise knowledge. In this respect, stochastic dominance was connected to imprecise probabilities by Denoeux ([61]), who generalised this notion to the comparison of belief functions ([187]). He proposed four extensions of stochastic dominance based on the orders between real intervals given in [78]. One step forward was made by Aiche and Dubois ([1]), by using stochastic dominance to compare random intervals stemming from rankings between real intervals, in a similar manner as Denoeux, and also in the comparison of fuzzy random variables ([105]).

On the other hand, the comparison of sets of random variables appears naturally in decision making under imprecision. In this sense, the usual *utility order* has already been extended in several ways to the comparison of sets of random variables: interval dominance ([219]), maximax ([184]) and maximin criteria ([82]), and E-admissibility ([107]). See a survey on this topic in ([202]).

With respect to statistical preference, Couso and Sánchez ([46]) proposed it as a method for comparing sets of desirable gambles (see [205, Sec. 2.2.4] for further information). Also, Couso and Dubois ([43]) proposed a common formulation for both statistical

preference and stochastic dominance to the comparison of imprecise probability models, and they studied its formulation in terms of expected utility.

Our aim here is to consider a more general situation: we start from a binary relation, that may be stochastic dominance, statistical preference or any other, as in Section 2.1, and extend it to the comparison of sets of random variables. We shall consider six possible extensions of the binary relation, and we shall study the connections between them. Afterwards, we consider the particular cases when the binary relation is stochastic dominance or statistical preference. As we shall see, our approach is more general than that of Denoeux, since the comparison of belief functions arises a particular case. On the other hand, our approach differs from the one of [43, 46] because they considered the comparison of sets of desirable gambles instead of sets of random variables, and the underlying philosophy of their approach is slightly different to ours.

After these general considerations, we shall focus on two scenarios that can be embedded into the comparison of sets of random variables: the comparison of two alternatives with imprecision either in the utilities or in the beliefs. The former will be formulated by means of random sets, and their comparison will be made by means of the associated sets of measurable selections. In the latter, we shall assume that there is a set of probability measures modelling the real probability measure of the probability space.

Since there could be imprecision on the initial probability, we devote the next section to the modelling of the joint distribution in an imprecise framework. For this aim, we shall investigate how the bivariate distribution can be expressed when there is imprecision in the initial probability. Then, we investigate bivariate p-boxes, and in particular how sets of bivariate distribution functions can define a bivariate p-box, and we study if it is possible to formulate an imprecise version of the famous Sklar's Theorem (see Theorem 2.27).

We conclude the chapter with several applications. First of all, we use imprecise stochastic dominance to compare sets of Lorenz Curves and cancer survival rates. Secondly, we use a multi criteria decision making problem to illustrate how imprecise stochastic orders can be applied in a context of imprecision either in the utilities or in the beliefs.

4.1 generalisation of the binary relations to the comparison of sets of random variables

In the following, we propose a number of methods for comparing pairs of sets of variables which are based on performing pairwise comparisons of elements within these sets. First we shall give our definitions for the case where the comparisons of the elements are made by means of a binary relation, as we did at the beginning of Section 2.1, and later we

shall apply them to the particular cases where this binary relation consists of stochastic dominance or statistical preference.

We shall consider a probability space (Ω, \mathcal{A}, P) and an ordered utility scale Ω' , that in some situations will be considered as numerical. We shall also consider sets of random variables, defined from the probability space to Ω' , that will be denoted by $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \dots$

We begin with the extension of a binary relation to the comparison of sets of random variables.

Definition 4.1 *Let \succeq be a binary relation between random variables defined from a probability space (Ω, \mathcal{A}, P) to an ordered utility scale Ω' . Given two sets of random variables \mathcal{X} and \mathcal{Y} , we say that:*

1. $\mathcal{X} \succeq_1 \mathcal{Y}$ if and only if for every $X \in \mathcal{X}, Y \in \mathcal{Y}$ it holds that $X \succeq Y$.
2. $\mathcal{X} \succeq_2 \mathcal{Y}$ if and only if there is some $X \in \mathcal{X}$ such that $X \succeq Y$ for every $Y \in \mathcal{Y}$.
3. $\mathcal{X} \succeq_3 \mathcal{Y}$ if and only if for every $Y \in \mathcal{Y}$ there is some $X \in \mathcal{X}$ such that $X \succeq Y$.
4. $\mathcal{X} \succeq_4 \mathcal{Y}$ if and only if there are $X \in \mathcal{X}, Y \in \mathcal{Y}$ such that $X \succeq Y$.
5. $\mathcal{X} \succeq_5 \mathcal{Y}$ if and only if there is some $Y \in \mathcal{Y}$ such that $X \succeq Y$ for every $X \in \mathcal{X}$.
6. $\mathcal{X} \succeq_6 \mathcal{Y}$ if and only if for every $X \in \mathcal{X}$ there is $Y \in \mathcal{Y}$ such that $X \succeq Y$.

Remark 4.2 *As we did in Definition 2.1, from any of these definitions we can infer immediately a relation of strict preference (\succ_i) and the indifference (\equiv_i):*

$$\begin{aligned} \mathcal{X} \succ_i \mathcal{Y} &\Leftrightarrow \mathcal{X} \succeq_i \mathcal{Y} \text{ and } \mathcal{Y} \not\succeq_i \mathcal{X}, \\ \mathcal{X} \equiv_i \mathcal{Y} &\Leftrightarrow \mathcal{X} \succeq_i \mathcal{Y} \text{ and } \mathcal{Y} \succeq_i \mathcal{X}, \end{aligned}$$

for any $i = 1, \dots, 6$. Moreover, we say that \mathcal{X} and \mathcal{Y} are incomparable with respect to \succeq_i when $\mathcal{X} \not\succeq_i \mathcal{Y}$ and $\mathcal{Y} \not\succeq_i \mathcal{X}$. ♦

The conditions in this definition can be given the following interpretation. \succeq_1 means that any alternative in \mathcal{X} is \succeq -preferred to any alternative in \mathcal{Y} , and as such it is related to the idea of interval dominance from decision making with sets of probabilities [219]. Conditions \succeq_2 and \succeq_3 mean that the “best” alternative in \mathcal{X} is \succeq -better than the “best” alternative in \mathcal{Y} . The difference between them lies in whether there is a maximal element in \mathcal{X} in the order determined by \succeq . These two conditions are related to the Γ -maximax criteria considered in [184]. On the other hand, conditions \succeq_5 and \succeq_6 mean that the “worst” alternative in \mathcal{X} is \succeq -preferred to the “worst” alternative in \mathcal{Y} , and are related to the Γ -maximin criteria in [20, 82]. Again, the difference between them lies in whether there is a minimum element in \mathcal{Y} with respect to the order determined by \succeq or not.

Finally, \succeq_4 is a weakened version of \succeq_1 , in the sense that it only requires that some alternative in \mathcal{X} is \succeq -preferred to some other alternative in \mathcal{Y} , instead of requiring it for any pair in \mathcal{X}, \mathcal{Y} .

Taking this interpretation into account, it is not difficult to establish the following relationships between the definitions.

Proposition 4.3 *The following implications hold:*

- $\succeq_1 \Rightarrow \succeq_2 \Rightarrow \succeq_3 \Rightarrow \succeq_4$.
- $\succeq_1 \Rightarrow \succeq_5 \Rightarrow \succeq_6 \Rightarrow \succeq_4$.

Proof: ($\succeq_1 \Rightarrow \succeq_2$): If $X \succeq Y$ for every $X \in \mathcal{X}, Y \in \mathcal{Y}$, in particular given any $X \in \mathcal{X}$ it holds that $X \succeq Y$ for every $Y \in \mathcal{Y}$.

($\succeq_2 \Rightarrow \succeq_3$): If there exists $X \in \mathcal{X}$ such that $X \succeq Y$ for every $Y \in \mathcal{Y}$, the condition in \succeq_3 is satisfied with respect to X for every $Y \in \mathcal{Y}$.

($\succeq_3 \Rightarrow \succeq_4$): If for every $Y \in \mathcal{Y}$ there exists $X_Y \in \mathcal{X}$ such that $X_Y \succeq Y$, we have a pair $(X_Y, Y) \in \mathcal{X} \times \mathcal{Y}$ such that $X_Y \succeq Y$.

($\succeq_1 \Rightarrow \succeq_5$): If $X \succeq Y$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, in particular given any $Y \in \mathcal{Y}$ it holds that $X \succeq Y$ for every $X \in \mathcal{X}$.

($\succeq_5 \Rightarrow \succeq_6$): If there is some $Y \in \mathcal{Y}$ such that $X \succeq Y$ for every $X \in \mathcal{X}$, in particular, for every $X \in \mathcal{X}$ it holds that $X \succeq Y$.

($\succeq_6 \Rightarrow \succeq_4$): If for every $X \in \mathcal{X}$ there exists $Y_X \in \mathcal{Y}$ such that $X \succeq Y_X$, we have a pair $(X, Y_X) \in \mathcal{X} \times \mathcal{Y}$ such that $X \succeq Y_X$. ■

The previous implications are depicted in Figure 4.1. Other relationships between the six definitions do not hold in general, as we can see in the following example.

Example 4.4 *Consider a probability space with only one element ω , and let δ_x denote the random variable satisfying $\delta_x(\omega) = x$. Consider also the binary relation \succeq such that:*

$$X \succeq Y \Leftrightarrow X(\omega) \geq Y(\omega). \quad (4.1)$$

If we take $\mathcal{X} = \{\delta_1, \delta_3\}$ and $\mathcal{Y} = \{\delta_2\}$, it follows that $\delta_3 \succ \delta_2 \succ \delta_1$, whence, applying Definition 4.1, we have that:

$$\mathcal{X} \succ_2 \mathcal{Y}, \quad \mathcal{X} \succ_3 \mathcal{Y}, \quad \mathcal{X} \equiv_4 \mathcal{Y}, \quad \mathcal{Y} \succ_5 \mathcal{X}, \quad \mathcal{Y} \succ_6 \mathcal{X}$$

and \mathcal{X} and \mathcal{Y} are incomparable with respect to the first extension. From this we deduce that:

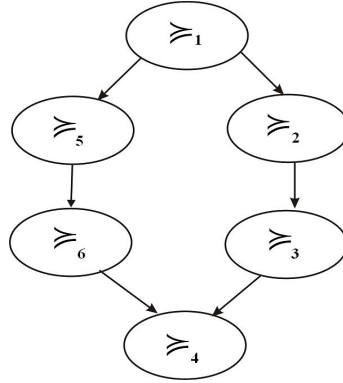


Figure 4.1: Relationships among the different extensions of the binary relation for the comparison of sets of random variables.

- $\succeq_2 \not\approx, \succeq_1, \succeq_5, \succeq_6$ and therefore $\succeq_3 \not\approx, \succeq_1, \succeq_5, \succeq_6$.
- $\succeq_4 \not\approx, \succeq_1, \succeq_2, \succeq_3, \succeq_5, \succeq_6$.
- $\succeq_5 \not\approx, \succeq_1, \succeq_2, \succeq_3$ and therefore $\succeq_6 \not\approx, \succeq_1, \succeq_2, \succeq_3$.

Next, given $\mathcal{X} = \mathcal{Y} = \{\delta_x : x \in (0, 1)\}$, we have that $\mathcal{X} \equiv_3 \mathcal{Y}$ and $\mathcal{X} \equiv_6 \mathcal{Y}$, because $\delta_x \equiv \delta_x$ for all $x \in (0, 1)$. However, \mathcal{X} and \mathcal{Y} are incomparable with respect to second and fifth definitions, because there are not $x_1, x_2 \in (0, 1)$ for which $\delta_{x_1} \succeq \delta_r$ and $\delta_r \succeq \delta_{x_2}$ for all $r \in (0, 1)$. Hence:

- $\succeq_3 \not\approx \succeq_2$.
- $\succeq_6 \not\approx \succeq_5$.♦

Remark 4.5 In some cases, it may be interesting to combine some of these definitions, for instance to consider \mathcal{X} preferred to \mathcal{Y} when it is preferred according to definitions \succeq_2 and \succeq_5 . Taking into account the implications depicted in Proposition 4.3, the combinations that produce new conditions are those where we take one condition out of $\{\succeq_2, \succeq_3\}$ together with one out of $\{\succeq_5, \succeq_6\}$.

If we combine for instance \succeq_2 with \succeq_5 , we can introduce the extension, denoted by $\succeq_{2,5}$, and defined by:

$$\mathcal{X} \succeq_{2,5} \mathcal{Y} \Leftrightarrow \mathcal{X} \succeq_2 \mathcal{Y} \text{ and } \mathcal{X} \succeq_5 \mathcal{Y}.$$

Then, $\succeq_{2,5}$ requires that \mathcal{X} has a \succeq -best case scenario which is better than any situation in \mathcal{Y} and that \mathcal{Y} has a \succeq -worst case which is worse than any situation in \mathcal{X} . This turns

out to be an intermediate condition between \succeq_1 and each of \succeq_2 and \succeq_5 , and it can be derived from the previous example that it is not equivalent to any of them. \blacklozenge

The implications in Proposition 4.3 can also be seen easily in the case where \mathcal{X} and \mathcal{Y} are finite sets, $\mathcal{X} = \{X_1, \dots, X_n\}$ and $\mathcal{Y} = \{Y_1, \dots, Y_m\}$. Then if we denote by M the $n \times m$ matrix where

$$M_{i,j} = \begin{cases} 1 & \text{if } X_i \succeq Y_j \\ 0 & \text{otherwise,} \end{cases}$$

the above definitions are characterised in the following way:

- $\mathcal{X} \succeq_1 \mathcal{Y} \Leftrightarrow M = 1_{n,m}$.
- $\mathcal{X} \succeq_2 \mathcal{Y} \Leftrightarrow \exists i \in \{1, \dots, n\}$ such that $M_{i,\cdot} = 1_{1,m}$.
- $\mathcal{X} \succeq_3 \mathcal{Y} \Leftrightarrow \exists j \in \{1, \dots, m\}$ such that $M_{\cdot,j} = 0_{n,1}$.
- $\mathcal{X} \succeq_4 \mathcal{Y} \Leftrightarrow M \neq 0_{n,m}$.
- $\mathcal{X} \succeq_5 \mathcal{Y} \Leftrightarrow \exists j \in \{1, \dots, m\}$ such that $M_{\cdot,j} = 1_{n,1}$.
- $\mathcal{X} \succeq_6 \mathcal{Y} \Leftrightarrow \exists i \in \{1, \dots, n\}$ such that $M_{i,\cdot} = 0_{1,m}$.

Observe that, as we have already seen, for any binary relation \succeq , its extensions \succeq_2 and \succeq_3 (respectively \succeq_5 and \succeq_6) are quite related: both compare the best (respectively, the worst) alternatives within each set \mathcal{X}, \mathcal{Y} . Since the difference between them lies on whether there is a maximal (respectively, minimal) element within each of these sets or not, we can easily give a necessary and sufficient condition for the equivalences $\succeq_2 \Leftrightarrow \succeq_3$ and $\succeq_5 \Leftrightarrow \succeq_6$.

Proposition 4.6 *Let \succeq be a binary relation on the set of random variables that is reflexive and transitive.*

- (a) *Given a set \mathcal{X} of random variables, $\mathcal{X} \succeq_3 \mathcal{Y} \Rightarrow \mathcal{X} \succeq_2 \mathcal{Y}$ for any set of variables \mathcal{Y} if and only if \mathcal{X} has a maximum element under \succeq .*
- (b) *Given a set \mathcal{Y} of random variables, $\mathcal{X} \succeq_6 \mathcal{Y} \Rightarrow \mathcal{X} \succeq_5 \mathcal{Y}$ for any set of variables \mathcal{X} if and only if \mathcal{Y} has a minimum element under \succeq .*

Proof:

- (a) Assume that \mathcal{X} has a maximum element X such that $X \succeq X'$ for every $X' \in \mathcal{X}$. If $\mathcal{X} \succeq_3 \mathcal{Y}$, then for every $Y \in \mathcal{Y}$ there is some $X_Y \in \mathcal{X}$ such that $X_Y \succeq Y$. Since

\succeq is transitive, we deduce that $X \succeq X_Y \succeq Y$, and then $X \succeq Y$ for every $Y \in \mathcal{Y}$, and as a consequence $\mathcal{X} \succeq_2 \mathcal{Y}$.

Conversely, if \mathcal{X} does not have a maximum element, we can take $\mathcal{Y} = \mathcal{X}$ and we would have $\mathcal{X} \equiv_3 \mathcal{Y}$ because \succeq is reflexive; however, \mathcal{X} and \mathcal{Y} are incomparable with respect to \succeq_2 because \mathcal{X} does not have a maximum element.

- (b) Similarly, if \mathcal{Y} has a minimum element Y , it holds that $Y' \succeq Y$ for any $Y' \in \mathcal{Y}$. If $\mathcal{X} \succeq_6 \mathcal{Y}$, then for every $X \in \mathcal{X}$ there exists $Y_X \in \mathcal{Y}$ such that $X \succeq Y_X$, and since \succeq is transitive we obtain that $X \succeq Y$ for every $X \in \mathcal{X}$, whence $\mathcal{X} \succeq_5 \mathcal{Y}$.

Conversely, if \mathcal{Y} does not have a minimum element, we can take $\mathcal{X} = \mathcal{Y}$ and we would have $\mathcal{X} \equiv_6 \mathcal{Y}$ because \succeq is reflexive; however, \mathcal{X} and \mathcal{Y} are incomparable with respect to \succeq_5 because \mathcal{Y} does not have a minimum element. ■

Under some conditions, we can also give a simpler characterisation of the above properties:

Proposition 4.7 *Let \succeq be a binary relation between random variables, and assume that it satisfies the Pareto Dominance condition:*

$$X(\omega) \geq Y(\omega) \quad \forall \omega \Rightarrow X \succeq Y. \quad (4.2)$$

Consider two sets of random variables \mathcal{X}, \mathcal{Y} . If the random variables $\min \mathcal{X}$, $\max \mathcal{X}$ exist and belong to \mathcal{X} and $\min \mathcal{Y}$, $\max \mathcal{Y}$ exist and belong to \mathcal{Y} , then:

- (a) $\mathcal{X} \succeq_1 \mathcal{Y} \Leftrightarrow \min \mathcal{X} \succeq \max \mathcal{Y}$.
- (b) $\mathcal{X} \succeq_2 \mathcal{Y} \Leftrightarrow \mathcal{X} \succeq_3 \mathcal{Y} \Leftrightarrow \max \mathcal{X} \succeq \max \mathcal{Y}$.
- (c) $\mathcal{X} \succeq_4 \mathcal{Y} \Leftrightarrow \max \mathcal{X} \succeq \min \mathcal{Y}$.
- (d) $\mathcal{X} \succeq_5 \mathcal{Y} \Leftrightarrow \mathcal{X} \succeq_6 \mathcal{Y} \Leftrightarrow \min \mathcal{X} \succeq \min \mathcal{Y}$.

Proof: Note that when both \mathcal{X}, \mathcal{Y} include a maximum and a minimum random variable, Equation (4.2) implies that for every $X \in \mathcal{X}, Y \in \mathcal{Y}$,

$$\min \mathcal{X} \succeq Y \Rightarrow X \succeq Y \Rightarrow \max \mathcal{X} \succeq Y$$

and

$$X \succeq \max \mathcal{Y} \Rightarrow X \succeq Y \Rightarrow X \succeq \min \mathcal{Y}.$$

Then:

- (a) Since $\min \mathcal{X} \succeq \max \mathcal{Y}$, it is obvious that $\mathcal{X} \succeq_1 \mathcal{Y}$. On the other hand, using the previous equations, if every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ satisfy $X \succeq Y$, then also $\min \mathcal{X} \succeq \max \mathcal{Y}$.

- (c) Since $\max \mathcal{X} \succeq \min \mathcal{Y}$, and $\max \mathcal{X} \in \mathcal{X}$ and $\min \mathcal{Y} \in \mathcal{Y}$, then $\mathcal{X} \succeq_4 \mathcal{Y}$. On the other hand, using the previous equations, if $X \succeq Y$ for some $X \in \mathcal{X}, Y \in \mathcal{Y}$, also $\max \mathcal{X} \succeq \min \mathcal{Y}$.
- (b,d) Using the previous equations, \mathcal{X} has a maximum element and \mathcal{Y} has a minimum element under \succeq . By Proposition 4.6, $\mathcal{X} \succeq_3 \mathcal{Y} \Leftrightarrow \mathcal{X} \succeq_2 \mathcal{Y}$ and $\mathcal{X} \succeq_6 \mathcal{Y} \Leftrightarrow \mathcal{X} \succeq_5 \mathcal{Y}$. The remaining equivalence can be established in an analogous manner to the previous cases. ■

Remark 4.8 According to Remark 4.5, under the conditions of the previous result, it is immediate that $\mathcal{X} \succeq_{2,5} \mathcal{Y}$ if and only if $\max \mathcal{X} \succeq \max \mathcal{Y}$ and $\min \mathcal{X} \succeq \min \mathcal{Y}$. ♦

Next we investigate which properties of the binary relation \succeq hold onto the extensions $\succeq_1, \dots, \succeq_6$. Obviously, since all these definitions become \succeq in the case of singletons, if \succeq is not reflexive (resp., antisymmetric, transitive), neither are \succeq_i , for $i = 1, \dots, 6$. Conversely, we can establish the following result.

Proposition 4.9 Let \succeq be a binary relation on random variables, and let $\succeq_i, i = 1, \dots, 6$ be its extensions to sets of random variables, given by Definition 4.1.

- (a) If \succeq is reflexive, so are \succeq_3, \succeq_4 and \succeq_6 .
- (b) If \succeq is antisymmetric, so is \succeq_1 .
- (c) If \succeq is transitive, so are \succeq_i for $i = 1, 2, 3, 5, 6$.

Proof: First of all, if \succeq is reflexive, $X \equiv X$ for any random variable X , and applying Definition 4.1 we deduce that $\mathcal{X} \succeq_i \mathcal{X}$ for any $i = 3, 4, 6$ and any set of random variables \mathcal{X} .

Secondly, assume that \succeq is antisymmetric and that two sets of random variables \mathcal{X}, \mathcal{Y} satisfy $\mathcal{X} \succeq_1 \mathcal{Y}$ and $\mathcal{Y} \succeq_1 \mathcal{X}$. Then, $X \succeq Y$ and $Y \succeq X$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, and by the antisymmetry property of \succeq , we deduce that $X = Y$ for every $X \in \mathcal{X}, Y \in \mathcal{Y}$. But this can only be if $\mathcal{X} = \{Z\} = \mathcal{Y}$ for some random variable Z . As a consequence, \succeq_1 is antisymmetric.

Finally, assume that \succeq is transitive, and let us show that so are \succeq_i for $i = 1, 2, 3, 5, 6$. Consider three sets of random variables $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$:

1. If $\mathcal{X} \succeq_1 \mathcal{Y}$ and $\mathcal{Y} \succeq_1 \mathcal{Z}$ then $X \succeq Y$ and $Y \succeq Z$ for every $X \in \mathcal{X}, Y \in \mathcal{Y}, Z \in \mathcal{Z}$. Applying the transitivity of \succeq , we deduce that $X \succeq Z$ for every $X \in \mathcal{X}, Z \in \mathcal{Z}$, and as a consequence $\mathcal{X} \succeq_1 \mathcal{Z}$.

2. If $\mathcal{X} \succeq_2 \mathcal{Y}$ and $\mathcal{Y} \succeq_2 \mathcal{Z}$, there is $X \in \mathcal{X}$ such that $X \succeq Y$ for every $Y \in \mathcal{Y}$ and there is $Y^* \in \mathcal{Y}$ such that $Y^* \succeq Z$ for every $Z \in \mathcal{Z}$. In particular, $X \succeq Y^* \succeq Z$ for every $Z \in \mathcal{Z}$, whence, by the transitivity of \succeq , $\mathcal{X} \succeq_2 \mathcal{Z}$.
3. If $\mathcal{X} \succeq_3 \mathcal{Y}$ and $\mathcal{Y} \succeq_3 \mathcal{Z}$, for every $Y \in \mathcal{Y}$ there is some $X_Y \in \mathcal{X}$ such that $X_Y \succeq Y$, and for every $Z \in \mathcal{Z}$ there is $Y_Z \in \mathcal{Y}$ such that $Y_Z \succeq Z$. As a consequence, for every $Z \in \mathcal{Z}$ it holds that $X_{Y_Z} \succeq Z$, and therefore $\mathcal{X} \succeq_3 \mathcal{Z}$.

The proof of the transitivity of \succeq_5 and \succeq_6 holds by analogy to that of \succeq_2 and \succeq_3 , respectively. \blacksquare

Our next example shows that reflexivity and antisymmetry do not hold for definitions different than the ones of statements (a) and (b). To show that the fourth extension is not transitive in general, even when the binary relation is, we refer to Example 4.18, where we shall show that the fourth extension is not transitive when considering the binary relation \succeq to be the first degree stochastic dominance.

Example 4.10 Consider the universe $\Omega = \{\omega\}$ and, as we made in Example 4.4, denote by δ_x the random variable such that $\delta_x(\omega) = x$, and the binary relation defined in Equation (4.1). Consider the set of random variables \mathcal{X} defined by $\mathcal{X} = \{\delta_x : x \in (0, 1)\}$. Then, although \succeq is reflexive, \mathcal{X} is incomparable with itself with respect to \succeq_1, \succeq_2 and \succeq_5 . Now, consider the sets of random variables \mathcal{X} and \mathcal{Y} defined by:

$$\mathcal{X} = \{\delta_x : x \in [0, 1]\} \text{ and } \mathcal{Y} = \{\delta_x : x \in [0, 1] \setminus \{0.5\}\}.$$

Then, $\mathcal{X} \equiv_i \mathcal{Y}$ for any $i = 2, 3, 4, 5, 6$, but $\mathcal{X} \neq \mathcal{Y}$, while \succeq is an antisymmetric relation. \blacklozenge

Another interesting property in a binary relation is that of *completeness*, which means that given any two elements, either one is preferred to the other or they are indifferent, but they are never incomparable. From Proposition 4.3, it follows that the incomparable pairs with respect to an extension \succeq_i are also incomparable with respect to the stronger extensions. The following result shows that if \succeq is a complete relation, then its weakest extensions (namely, \succeq_3, \succeq_4 and \succeq_6) also induce complete binary relations:

Proposition 4.11 Consider a binary relation \succeq between random variables, and let \succeq_i , for $i = 1, \dots, 6$, be its extensions to sets of random variables given by Definition 4.1. If \succeq is complete, then so are \succeq_3, \succeq_4 and \succeq_6 .

Proof: Let \mathcal{X}, \mathcal{Y} be two sets of random variables, and assume that $\mathcal{X} \not\succeq_3 \mathcal{Y}$. Then there is some $Y \in \mathcal{Y}$ such that $X \not\succeq Y$ for all $X \in \mathcal{X}$. But since \succeq is a complete relation, this means that $Y \succeq X$ for all $X \in \mathcal{X}$. As a consequence, $\mathcal{Y} \succeq_2 \mathcal{X}$, and applying Proposition 4.3 we deduce that $\mathcal{Y} \succeq_3 \mathcal{X}$. Hence, the binary relation \succeq_3 is complete.

\succeq	\succeq_1	\succeq_2	\succeq_3	\succeq_4	\succeq_5	\succeq_6
Reflexive			•	•		•
Antisymmetric	•					
Transitive	•	•	•		•	•
Complete			•	•		•

Table 4.1: Summary of the properties of the binary relation \succeq that hold onto their extensions $\succeq_1, \dots, \succeq_6$.

On the other hand, if $\mathcal{X} \not\succeq_4 \mathcal{Y}$, we deduce from Proposition 4.3 that also $\mathcal{X} \not\succeq_3 \mathcal{Y}$, whence the above reasoning implies that $\mathcal{Y} \succeq_3 \mathcal{X}$ and again from Proposition 4.3 we deduce that $\mathcal{Y} \succeq_4 \mathcal{X}$.

The proof that \succeq_6 also induces a complete relation is analogous. ■

Let us now give an example where we see that the completeness of the binary relationship does not imply the completeness of the extensions $\succeq_1, \succeq_2, \succeq_5$.

Example 4.12 Consider again Example 4.10, and take the sets of random variables $\mathcal{X} = \mathcal{Y} = \{\delta_x : x \in (0, 1)\}$ and the binary relation \succeq defined in Equation (4.1). Although \succeq is complete, \mathcal{X} and \mathcal{Y} are incomparable with respect to \succeq_1, \succeq_2 and \succeq_5 . ♦

Table 4.1 summarises the properties we have investigated in Propositions 4.9 and 4.11.

Remark 4.13 Although in this report we shall focus on the particular application of Definition 4.1 to the relation \succeq associated with stochastic dominance or statistical preference, there are other cases of interest. Perhaps the most important one is that where the comparison between pairs of random variables is made by means of their expected utility:

$$X \succeq Y \Leftrightarrow E(X) \geq E(Y);$$

it is not difficult to see that Definition 4.1 gives rise to some well-known generalisations of expected utility that are formulated in terms of lower and upper expectations. Consider two sets \mathcal{X}, \mathcal{Y} and assume that the expectations of all their elements exist. Then with respect to definition \succeq_1 it holds that:

$$\mathcal{X} \succeq_1 \mathcal{Y} \Leftrightarrow \underline{E}(\mathcal{X}) = \inf_{X \in \mathcal{X}} E(X) \geq \sup_{Y \in \mathcal{Y}} E(Y) = \overline{E}(\mathcal{Y}),$$

which relates this notion to the concept of interval dominance considered in [219].

If we now consider definition \succeq_3 , it holds that

$$\mathcal{X} \succeq_3 \mathcal{Y} \Rightarrow \overline{E}(\mathcal{X}) = \sup_{X \in \mathcal{X}} E(X) \geq \sup_{Y \in \mathcal{Y}} E(Y) = \overline{E}(\mathcal{Y}).$$

Thus, definition \succeq_3 is stronger than the maximax criterium [184], which is based on comparing the best possibilities in our sets of alternatives. Similarly, if we consider definition \succeq_6 it holds that:

$$\mathcal{X} \succeq_6 \mathcal{Y} \Rightarrow \underline{E}(\mathcal{X}) = \inf_{X \in \mathcal{X}} E(X) \geq \inf_{Y \in \mathcal{Y}} E(Y) = \underline{E}(\mathcal{Y}).$$

Thus, definition \succeq_6 is stronger than the maximin criterium [82], which compares the worst possibilities within the sets of alternatives.

Finally, definition \succeq_4 implies that

$$\mathcal{X} \succeq_4 \mathcal{Y} \Rightarrow \overline{E}(\mathcal{X}) = \sup_{X \in \mathcal{X}} E(X) \geq \inf_{Y \in \mathcal{Y}} E(Y) = \underline{E}(\mathcal{Y}),$$

so if \mathcal{X} is \succeq_4 -preferred to \mathcal{Y} then it is also preferred with respect to the criterion of E-admissibility from [107]. See [43, 202] for related comments. \blacklozenge

4.1.1 Imprecise stochastic dominance

In this subsection, we explore in some detail the case where the binary relation \succeq is the one associated with the notion of first degree stochastic dominance we have introduced in Definition 2.2, i.e., the relation \succeq is defined by \succeq_{FSD} . We call this extension imprecise stochastic dominance. We shall assume that the utility space Ω' is $[0, 1]$, although the results can be immediately extended to any bounded interval of real numbers. Since stochastic dominance is based on the comparison of cumulative distribution functions associated with the random variables, we shall employ the notation $F_X \succeq_{\text{FSD}} F_Y$ instead of $X \succeq_{\text{FSD}} Y$. For the same reason, along this subsection we will consider sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y instead of sets of random variables \mathcal{X} and \mathcal{Y} .

Remark 4.14 From now on, we shall say that a set of distribution functions \mathcal{F}_X is (FSD_i) -preferred or that it (FSD_i) -stochastically dominates another set of distribution functions \mathcal{F}_Y when $\mathcal{F}_X \succeq_{\text{FSD}_i} \mathcal{F}_Y$. We will also use the notation $\succeq_{\text{FSD}_{i,j}}$ when both \succeq_{FSD_i} and \succeq_{FSD_j} hold. \blacklozenge

An illustration of the six extensions of Definition 4.1 when considering stochastic dominance is given in Figure 4.2, where we compare the set of distribution functions represented by a continuous line (that we shall call continuous distributions in this paragraph) with the set of distribution functions represented by a dotted line (that we shall call dotted distributions). On the one hand, in the left picture the set of continuous distributions (FSD_1) -stochastically dominates the set of dotted distributions. In the right picture, there is a continuous distribution that dominates all dotted distributions, and a dotted distribution which is dominated by all continuous distributions. This means

that the set of continuous distributions stochastically dominates the set of dotted distributions with respect to the second to sixth definitions. Since there is also a dotted distribution that is dominated by a continuous distribution, we deduce that the set of continuous distributions and the set of dotted distributions are equivalent with respect to the fourth definition. Notice that the binary relationship considered in Example 4.4

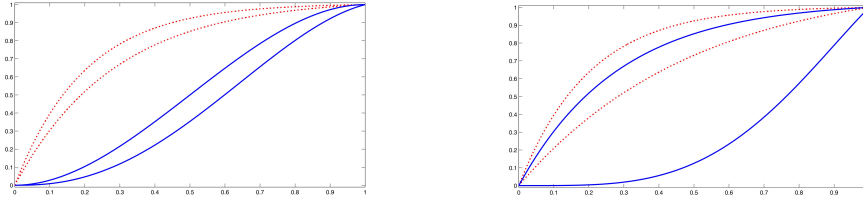


Figure 4.2: Examples of several definitions of imprecise stochastic dominance.

is equivalent to first degree stochastic dominance when the initial space Ω only has one element. Then, such example shows that the converse implications of Proposition 4.3 do not hold in general when considering the binary relation to be the first degree stochastic dominance.

Now, we investigate which properties hold when considering the strict imprecise stochastic dominance.

Proposition 4.15 *Consider the extensions of stochastic dominance given in Definition 4.1. It holds that:*

- $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succ_{\text{FSD}_3} \mathcal{F}_Y$.
- $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succ_{\text{FSD}_6} \mathcal{F}_Y$.

Proof: We begin proving that \succ_{FSD_2} implies \succ_{FSD_3} . Observe that $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$ is equivalent to:

- (I) $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Rightarrow \exists F_1^* \in \mathcal{F}_X$ such that $F_1^* \leq F_2$ for all $F_2 \in \mathcal{F}_Y$.
- (II) $\mathcal{F}_Y \not\prec_{\text{FSD}_2} \mathcal{F}_X \Rightarrow \forall F_2 \in \mathcal{F}_Y, \exists F_1 \in \mathcal{F}_X$ such that $F_2 \not\leq F_1$.

It follows from (I) and Proposition 4.3 that $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y$. We only have to prove that $\mathcal{F}_Y \not\prec_{\text{FSD}_3} \mathcal{F}_X$, or equivalently, that there is $F_1 \in \mathcal{F}_X$ such that $F_2 \not\leq F_1$ for any $F_2 \in \mathcal{F}_Y$. If F_1^* satisfies this property, the proof is finished. If not, there is some $F_2^* \in \mathcal{F}_Y$ such that $F_2^* \leq F_1^*$, whence $F_1^* = F_2^*$. Applying (II), there exists some $F_1' \in \mathcal{F}_X$ such that $F_1^* \not\leq F_1'$, which means that $F_1'(t) < F_1^*(t)$ for some t . As a consequence, $F_1'(t) < F_2(t)$ for any $F_2 \in \mathcal{F}_Y$, whence $\mathcal{F}_Y \not\prec_{\text{FSD}_3} \mathcal{F}_X$. Hence, $\mathcal{F}_X \succ_{\text{FSD}_3} \mathcal{F}_Y$.

Let us now prove that $\succ_{\text{FSD}_5} \Rightarrow \succ_{\text{FSD}_6}$. Similarly to the previous case, $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$ is equivalent to:

- (I) $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Rightarrow \exists F_2^* \in \mathcal{F}_Y$ such that $F_1 \leq F_2^*$ for all $F_1 \in \mathcal{F}_X$.
- (II) $\mathcal{F}_Y \not\prec_{\text{FSD}_5} \mathcal{F}_X \Rightarrow \forall F_1 \in \mathcal{F}_X, \exists F_2 \in \mathcal{F}_Y$ such that $F_2 \not\leq F_1$.

It follows from (I) and Proposition 4.3 that $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y$. We only have to prove that $\mathcal{F}_Y \not\prec_{\text{FSD}_6} \mathcal{F}_X$, or equivalently, that there is $F_2 \in \mathcal{F}_Y$ such that $F_2 \not\leq F_1$ for any $F_1 \in \mathcal{F}_X$. If F_2^* satisfies this property, the proof is finished. If not, there exists $F_1^* \in \mathcal{F}_X$ such that $F_2^* \leq F_1^*$, and applying (I) we deduce that $F_1^* = F_2^* \in \mathcal{F}_X$. Applying (II) we deduce that there is some $F_2' \in \mathcal{F}_Y$ such that $F_2' \not\leq F_1^*$, whence there is some t such that $F_2'(t) > F_1^*(t) = F_2^*(t) \geq F_1(t)$ for every $F_1 \in \mathcal{F}_X$. Hence, $F_2' \not\leq F_1$ for any $F_1 \in \mathcal{F}_X$ and the property holds. \blacksquare

Furthermore, next example shows that there are no other relationships between the strict extensions of stochastic dominance.

Example 4.16 Consider the same conditions of Example 4.4: $\Omega = \{\omega\}$, δ_x is the random variable given by $\delta_x(\omega) = x$ and \succeq is given by Equation (4.1), that is equivalent to \succeq_{FSD} in this case.

Take the sets $\mathcal{X} = \{\delta_1\}$ and $\mathcal{Y} = \{\delta_0, \delta_1\}$. It holds that:

$$\mathcal{X} \succ_{\text{FSD}_1} \mathcal{Y} \text{ and } \mathcal{X} \succ_{\text{FSD}_6} \mathcal{Y},$$

but $\mathcal{X} \equiv_{\text{FSD}_2} \mathcal{Y}$ and $\mathcal{X} \equiv_{\text{FSD}_4} \mathcal{Y}$. Then, $\succ_{\text{FSD}_1} \not\Rightarrow \succ_{\text{FSD}_2}$ and $\succ_{\text{FSD}_6} \not\Rightarrow \succ_{\text{FSD}_4}$.

If we consider the sets $\mathcal{X} = \{\delta_0, \delta_1\}$ and $\mathcal{Y} = \{\delta_0\}$, it holds that:

$$\mathcal{X} \succ_{\text{FSD}_1} \mathcal{Y} \text{ and } \mathcal{X} \succ_{\text{FSD}_3} \mathcal{Y},$$

but $\mathcal{X} \equiv_{\text{FSD}_5} \mathcal{Y}$ and $\mathcal{X} \equiv_{\text{FSD}_4} \mathcal{Y}$. Then, $\succ_{\text{FSD}_1} \not\Rightarrow \succ_{\text{FSD}_5}$ and $\succ_{\text{FSD}_3} \not\Rightarrow \succ_{\text{FSD}_4}$. \blacklozenge

With respect to the other results, since \succeq_{FSD} is reflexive and transitive, we can apply Proposition 4.6 and characterise the equivalences between \succeq_{FSD_2} and \succeq_{FSD_3} , and also between \succeq_{FSD_5} and \succeq_{FSD_6} by means of the existence of a maximum and a minimum value in the sets $\mathcal{F}_X, \mathcal{F}_Y$ we want to compare. Moreover, we can deduce from Proposition 4.9 and Examples 4.10 and 4.12 that \succeq_{FSD_i} is reflexive for $i = 3, 4, 6$ and transitive for $i = 1, 2, 3, 5, 6$. On the other hand, since two different random variables may induce the same distribution function, \succeq_{FSD} is not antisymmetric. Nevertheless, if we are dealing with sets of cumulative distribution functions instead of sets of random variables, \succeq_{FSD} becomes antisymmetric. Next example shows that (FSD_4) is not transitive in general.

Remark 4.17 Through this subsection we shall present several examples showing that the propositions established cannot be improved, in the sense that the missing implications

do not hold in general. Some of these examples will consider distribution functions associated with probability measures with finite supports. To fix notation, given $\vec{a} = (a_1, \dots, a_n)$ such that $a_1 + \dots + a_n = 1$, and $\vec{t} = (t_1, \dots, t_n)$ with $t_1 \leq \dots \leq t_n$, the function $F_{\vec{a}, \vec{t}}$ corresponds to the cumulative distribution function of the probability measure $P_{\vec{a}, \vec{t}}$ satisfying $P_{\vec{a}, \vec{t}}(\{t_i\}) = a_i$ for $i = 1, \dots, n$. Indeed, the only continuous distribution function we shall consider is the identity $F = id$, defined by $F(x) = id(x) = x$ for any $x \in [0, 1]$. ♦

Example 4.18 Consider the three sets of cumulative distribution functions \mathcal{F}_X , \mathcal{F}_Y and \mathcal{F}_Z defined by:

$$\mathcal{F}_X = \{F_{(0.5, 0.5), (0, 1)}\}, \quad \mathcal{F}_Z = \{F\}, \quad \mathcal{F}_Y = \mathcal{F}_X \cup \mathcal{F}_Z.$$

Since both sets \mathcal{F}_X and \mathcal{F}_Z are included in \mathcal{F}_Y , Proposition 4.29 later on assures that $\mathcal{F}_X \equiv_{\text{FSD}_4} \mathcal{F}_Y$ and $\mathcal{F}_Y \equiv_{\text{FSD}_4} \mathcal{F}_Z$. However, \mathcal{F}_X and \mathcal{F}_Z are not comparable, since the distribution functions $F_{(0.5, 0.5), (0, 1)}$ and F are not comparable with respect to first degree stochastic dominance. ♦

Since \succeq_{FSD} also complies with Pareto dominance (Equation (4.2)), we deduce from Proposition 4.7 that when the sets \mathcal{F}_X and \mathcal{F}_Y to compare have both a maximum and a minimum element, we can easily characterise the conditions \succeq_{FSD_i} , $i = 1, \dots, 6$ by comparing these maximum and minimum elements only. Finally, note that, as we already mentioned in Example 2.3, \succeq_{FSD} is not a complete relation, and as a consequence, Proposition 4.11 is not applicable in this context.

As we remarked in Section 2.2.1, p-boxes are one model within the theory of imprecise probabilities. Stochastic dominance between sets of probabilities or cumulative distribution functions can be studied by means of a p-box representation. Given any set of cumulative distribution functions \mathcal{F} , it induces a p-box $(\underline{F}, \overline{F})$, as we saw in Equation (2.16):

$$\underline{F}(x) := \inf_{F \in \mathcal{F}} F(x), \quad \overline{F}(x) := \sup_{F \in \mathcal{F}} F(x).$$

Our next result relates the imprecise stochastic dominance for sets of cumulative distribution functions to their associated p-box representation.

Proposition 4.19 Let \mathcal{F}_X and \mathcal{F}_Y be two sets of cumulative distribution functions, and denote by $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ the p-boxes they induce by means of Equation (2.16). Then the following statements hold:

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \overline{F}_X \succeq_{\text{FSD}} \underline{F}_Y$.
2. $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Rightarrow \underline{F}_X \succeq_{\text{FSD}} \underline{F}_Y$.
3. $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y \Rightarrow \underline{F}_X \succeq_{\text{FSD}} \underline{F}_Y$.

4. $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \Rightarrow \underline{F}_X \succeq_{\text{FSD}} \overline{F}_Y$.
5. $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Rightarrow \overline{F}_X \succeq_{\text{FSD}} \overline{F}_Y$.
6. $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y \Rightarrow \overline{F}_X \succeq_{\text{FSD}} \overline{F}_Y$.

Proof:

- (1) Note that $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y$ if and only if $F_1 \leq F_2$ for every $F_1 \in \mathcal{F}_X, F_2 \in \mathcal{F}_Y$, and this is equivalent to $\overline{F}_X = \sup_{F_1 \in \mathcal{F}_X} F_1 \leq \inf_{F_2 \in \mathcal{F}_Y} F_2 = \underline{F}_Y$.
 - (3) By hypothesis, for every $F_2 \in \mathcal{F}_Y$ there is some $F_1 \in \mathcal{F}_X$ such that $F_1 \leq F_2$. As a consequence, $\underline{F}_X \leq F_2 \forall F_2 \in \mathcal{F}_Y \Rightarrow \underline{F}_X \leq \inf_{F_2 \in \mathcal{F}_Y} F_2 = \underline{F}_Y$.
 - (4) If there are $F_1 \in \mathcal{F}_X$ and $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$, then $\underline{F}_X \leq F_1 \leq F_2 \leq \overline{F}_Y$.
 - (6) If for every $F_1 \in \mathcal{F}_X$ there is some $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$, then it holds that $\overline{F}_X = \sup_{F_1 \in \mathcal{F}_X} F_1 \leq \sup_{F_2 \in \mathcal{F}_Y} F_2 = \overline{F}_Y$.
- (2,5) The second (resp. fifth) statement follows from the third (resp., sixth) and Proposition 4.3. ■

Next example shows that the converse implications in the second to sixth statements do not hold in general.

Example 4.20 Take $\mathcal{F}_X = \{F_{(0.3,0.7),(0,1)}, F_{(0.2,0.8),(0.2,0.3)}\}$, $\mathcal{F}_Y = \{F\}$. They are incomparable under any of the definitions but $\underline{F}_X \leq \underline{F}_Y = F = \overline{F}_Y \leq \overline{F}_X$, from which we deduce that the converse implications in Proposition 4.19 do not hold. ◆

As we mentioned after Definition 4.1, the difference between (FSD_2) and (FSD_3) lies on whether the set of distribution functions \mathcal{F}_X has a “best case”, i.e., a smallest distribution function; similarly, the difference between (FSD_5) and (FSD_6) lies on whether \mathcal{F}_Y has a greatest distribution function. Taking this into account, we can easily adapt the conditions of Proposition 4.6 towards imprecise stochastic dominance:

Proposition 4.21 Let \mathcal{F}_X and \mathcal{F}_Y be two sets of cumulative distribution functions.

1. $\underline{F}_X \in \mathcal{F}_X \Rightarrow [\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y]$.
2. $\overline{F}_Y \in \mathcal{F}_Y \Rightarrow [\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y]$.

Proof: To see the first statement, use that by Proposition 4.3 $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y$ implies $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y$. Moreover, $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y$ if and only if for every $F_2 \in \mathcal{F}_Y$ there is

$F_1 \in \mathcal{F}_X$ such that $F_1 \leq F_2$. In particular, since $\underline{F}_X \leq F_1$ for every $F_1 \in \mathcal{F}_X$, it holds that $\underline{F}_X \leq F_2$ for every $F_2 \in \mathcal{F}_Y$, and consequently, as $\underline{F}_X \in \mathcal{F}_X$, that $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y$.

The proof of the second statement is analogous. ■

When both the lower and upper distributions belong to the corresponding p-box, they can be used to characterise the preferences between them. In that case, the stochastic dominance between two sets of cumulative distribution functions can be characterised by means of the relationships of stochastic dominance between their lower and upper distribution functions.

Corollary 4.22 *Let $\mathcal{F}_X, \mathcal{F}_Y$ be two sets of cumulative distribution functions, and let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ be their associated p-boxes. If $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$, then*

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \overline{F}_X \leq \underline{F}_Y$.
2. $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y \Leftrightarrow \underline{F}_X \leq \underline{F}_Y$.
3. $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \Leftrightarrow \underline{F}_X \leq \overline{F}_Y$.
4. $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y \Leftrightarrow \overline{F}_X \leq \overline{F}_Y$.

Proof: The first item has already been showed in Proposition 4.19. The equivalences between $(\text{FSD}_2) - (\text{FSD}_3)$ and $(\text{FSD}_5) - (\text{FSD}_6)$ are given by Proposition 4.21. Also, the direct implications of second, third and fourth items are given by Proposition 4.19. Let us prove the converse implications:

- If $\underline{F}_Y \geq \underline{F}_X \in \mathcal{F}_X$, there is some $F_1 \in \mathcal{F}_X$ such that $F_1 \leq F_2$ for all $F_2 \in \mathcal{F}_Y$, and as a consequence $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y$.
- If $\underline{F}_X \leq \overline{F}_Y$, then there exist $F_1 \in \mathcal{F}_X$ and $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$, whence $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y$.
- If $\overline{F}_X \leq \overline{F}_Y$, then since $\overline{F}_Y \in \mathcal{F}_Y$ then there is some $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$ for every $F_1 \in \mathcal{F}_X$, because $F_X \leq \overline{F}_X$ for any $F_X \in \mathcal{F}_X$. ■

In Section 2.1.1 we established a characterisation of stochastic dominance in terms of expectations: Theorem 2.10 assures that given two random variables X and Y , $X \succeq_{\text{FSD}} Y$ if and only if $E(u(X)) \geq E(u(Y))$ for every increasing function u . When we compare sets of random variables, we must replace these expectations by lower and upper expectations. For any given set of distribution functions \mathcal{F} and any increasing function $u : [0, 1] \rightarrow \mathbb{R}$, we shall denote $\underline{E}_{\mathcal{F}}(u) := \inf_{F \in \mathcal{F}} E_{P_F}(u)$ and $\overline{E}_{\mathcal{F}}(u) := \sup_{F \in \mathcal{F}} E_{P_F}(u)$.

Theorem 4.23 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y , and let \mathcal{U} be the set of all increasing functions $u : [0, 1] \rightarrow \mathbb{R}$. The following statements hold:*

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \underline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
2. $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Rightarrow \overline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
3. $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y \Rightarrow \overline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
4. $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \Rightarrow \overline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
5. $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Rightarrow \underline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
6. $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y \Rightarrow \underline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.

Proof:

1. First of all, $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y$ if and only if for every $F_1 \in \mathcal{F}_X$ and $F_2 \in \mathcal{F}_Y$ $F_1 \succeq_{\text{FSD}} F_2$. This is equivalent to $E_{P_1}(u) \geq E_{P_2}(u)$, for every $u \in \mathcal{U}$, and every $F_1 \in \mathcal{F}_X$ and $F_2 \in \mathcal{F}_Y$, where P_i is the probability associated with F_i , for $i = 1, 2$, and this in turn is equivalent to

$$\underline{E}_{\mathcal{F}_X}(u) = \inf\{E_{P_F}(u) \mid F \in \mathcal{F}_X\} \geq \sup\{E_{P_F}(u) \mid F \in \mathcal{F}_Y\} = \overline{E}_{\mathcal{F}_Y}(u)$$

for every $u \in \mathcal{U}$.

3. If $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y$, then for every $F_2 \in \mathcal{F}_Y$ there is $F_1 \in \mathcal{F}_X$ such that $F_1 \leq F_2$. Equivalently, for every $F_2 \in \mathcal{F}_Y$ there is $F_1 \in \mathcal{F}_X$ such that $E_{P_1}(u) \geq E_{P_2}(u)$ for every $u \in \mathcal{U}$. Then given $u \in \mathcal{U}$ and $F_2 \in \mathcal{F}_Y$,

$$E_{P_2}(u) \leq \sup\{E_{P_F}(u) \mid F \in \mathcal{F}_X\} = \overline{E}_{\mathcal{F}_X}(u),$$

and consequently

$$\overline{E}_{\mathcal{F}_Y}(u) = \sup\{E_{P_F}(u) \mid F \in \mathcal{F}_Y\} \leq \overline{E}_{\mathcal{F}_X}(u).$$

2. The second statement follows from the third one and from Proposition 4.3.
4. Let us assume that $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y$. Then, by definition there are $F_1 \in \mathcal{F}_X$ and $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$, or equivalently, $E_{P_1}(u) \geq E_{P_2}(u)$ for every $u \in \mathcal{U}$. We deduce that

$$\begin{aligned} \overline{E}_{\mathcal{F}_X}(u) &= \sup\{E_{P_F}(u) \mid F \in \mathcal{F}_X\} \geq E_{P_1}(u) \\ &\geq E_{P_2}(u) \geq \inf\{E_{P_F}(u) \mid F \in \mathcal{F}_Y\} = \underline{E}_{\mathcal{F}_Y}(u). \end{aligned}$$

6. If $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y$, then for every $F_1 \in \mathcal{F}_X$ there is $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$. Equivalently, for every $F_1 \in \mathcal{F}_X$, $E_{P_1}(u) \geq E_{P_2}(u)$ for some $F_2 \in \mathcal{F}_Y$ and for every $u \in \mathcal{U}$. Thus, for every $F_1 \in \mathcal{F}_X$ and $u \in \mathcal{U}$,

$$E_{P_1}(u) \geq \inf\{E_{P_F}(u) \mid F \in \mathcal{F}_Y\},$$

and consequently

$$\underline{E}_{\mathcal{F}_X}(u) = \inf\{E_{P_F}(u) \mid F \in \mathcal{F}_X\} \geq \inf\{E_{P_F}(u) \mid F \in \mathcal{F}_Y\} = \underline{E}_{\mathcal{F}_Y}(u).$$

5. Finally, the fifth statement follows from the sixth and from Proposition 4.3. \blacksquare

Remark 4.24 *If we consider the extension of stochastic dominance $\succeq_{\text{FSD}_{3,6}}$, that is, $\mathcal{F}_X \succeq_{\text{FSD}_{3,6}} \mathcal{F}_Y$ if and only if $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y$ and $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y$, it holds that:*

$$\mathcal{F}_X \succeq_{\text{FSD}_{3,6}} \mathcal{F}_Y \Rightarrow \begin{cases} \underline{E}_X \succeq_{\text{FSD}} \underline{E}_Y \text{ and } \overline{F}_X \succeq_{\text{FSD}} \overline{F}_Y. \\ \overline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u) \text{ and } \underline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u) \forall u \in \mathcal{U}. \end{cases} \quad (4.3)$$

With a similar notation, we can consider $\succeq_{\text{FSD}_{2,5}}$, and it holds that $\mathcal{F}_X \succeq_{\text{FSD}_{2,5}} \mathcal{F}_Y$ implies $\mathcal{F}_X \succeq_{\text{FSD}_{3,6}} \mathcal{F}_Y$. Then, from the previous results we deduce that $\mathcal{F}_X \succeq_{\text{FSD}_{2,5}} \mathcal{F}_Y$ also implies the results of Equation (4.3). \blacklozenge

Taking into account Equation (2.6), the above implications hold in particular when we replace the set \mathcal{U} by the subset \mathcal{U}^* of increasing and bounded functions $u : [0, 1] \rightarrow \mathbb{R}$. This will be useful when comparing random sets by means of stochastic dominance in Section 4.2.1.

Remark 4.25 *Theorem 4.23 shows that the extensions of first degree stochastic dominance to sets of alternatives are related to the comparison of the lower and upper expectations they induce. Taking this idea into account, we may introduce alternative definitions by considering a convex combination of these lower and upper expectations, in a similar way to the Hurwicz criterion [96]:*

$$\mathcal{F}_X \succeq_{\text{FSD}_H} \mathcal{F}_Y \Leftrightarrow \lambda \underline{E}_{\mathcal{F}_X}(u) + (1 - \lambda) \overline{E}_{\mathcal{F}_X}(u) \geq \lambda \underline{E}_{\mathcal{F}_Y}(u) + (1 - \lambda) \overline{E}_{\mathcal{F}_Y}(u),$$

for all $u \in \mathcal{U}$, where $\lambda \in [0, 1]$ plays the role of a pessimistic index. It is not difficult to see that

$$\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_{2,5}} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_{3,6}} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_H} \mathcal{F}_Y$$

and that the converses do not hold. \blacklozenge

When the bounds of the p-boxes belong the sets of distribution functions, the implications on Theorem 4.23 become equivalences.

Corollary 4.26 *Let \mathcal{F}_X and \mathcal{F}_Y be two sets of cumulative distribution functions, and let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ be their associated p-boxes. If $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$, then:*

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \underline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
2. $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y \Leftrightarrow \overline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
3. $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \Leftrightarrow \overline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.
4. $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y \Leftrightarrow \underline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u)$ for every $u \in \mathcal{U}$.

Proof: The proof is based on the fact that, since $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$, then:

$$\begin{aligned} \underline{E}_{\mathcal{F}_X}(u) &= E_{\overline{F}_X}(u), & \overline{E}_{\mathcal{F}_X}(u) &= E_{\underline{F}_X}(u), \\ \underline{E}_{\mathcal{F}_Y}(u) &= E_{\overline{F}_Y}(u), & \overline{E}_{\mathcal{F}_Y}(u) &= E_{\underline{F}_Y}(u). \end{aligned}$$

Then, applying Corollary 4.22, the implications directly hold. ■

It is also possible to consider the n -th degree stochastic dominance, for $n \geq 2$ as the binary relation in Definition 4.1. In that case, we shall denote by $\succeq_{n\text{SD}_i}$ or by $(n\text{SD}_i)$ its extensions. With this relation, we can also state similar results to the ones established for first degree stochastic dominance. For instance, the following statements hold for imprecise n -th degree stochastic dominance:

- $\mathcal{F}_X \succ_{n\text{SD}_2} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succ_{n\text{SD}_3} \mathcal{F}_Y$ (the proof is analogous to that of Proposition 4.15).
- $\mathcal{F}_X \succ_{n\text{SD}_5} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succ_{n\text{SD}_6} \mathcal{F}_Y$ (the proof is analogous to that of Proposition 4.15).
- $\mathcal{F}_X \succeq_{n\text{SD}_i} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{m\text{SD}_i} \mathcal{F}_Y$ for any $n < m$ (see Equation (2.4)).

In addition, the connection of the comparison of sets of cumulative distribution functions with the associated p-boxes (Proposition 4.19) or with the associated lower and upper expectations (Theorem 4.23) can also be stated for the imprecise n -th degree stochastic dominance as follows:

Proposition 4.27 *Let \mathcal{F}_X and \mathcal{F}_Y be two sets of cumulative distribution functions, and denote by $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ the associated p-boxes. Denote by \mathcal{U}_n^* the set of bounded and increasing functions $u: \mathbb{R} \rightarrow \mathbb{R}$ that are n -monotone. Then it holds that:*

- $\mathcal{F}_X \succeq_{n\text{SD}_1} \mathcal{F}_Y$ holds if and only if $\overline{F}_X \succeq_{n\text{SD}_1} \underline{F}_Y$, and this is equivalent to

$$\underline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u)$$

for every $u \in \mathcal{U}_n^*$.

- $\mathcal{F}_X \succeq_{\text{nSD}_2} \mathcal{F}_Y$ implies:

$$\underline{F}_X \succeq_{\text{nSD}_2} \underline{F}_Y \text{ and } \overline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u) \text{ for every } u \in \mathcal{U}_n^*.$$

- $\mathcal{F}_X \succeq_{\text{nSD}_3} \mathcal{F}_Y$ implies:

$$\underline{F}_X \succeq_{\text{nSD}_3} \underline{F}_Y \text{ and } \overline{E}_{\mathcal{F}_X}(u) \geq \overline{E}_{\mathcal{F}_Y}(u) \text{ for every } u \in \mathcal{U}_n^*.$$

- $\mathcal{F}_X \succeq_{\text{nSD}_4} \mathcal{F}_Y$ implies:

$$\underline{F}_X \succeq_{\text{FSD}_4} \overline{F}_Y \text{ and } \overline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u) \text{ for every } u \in \mathcal{U}_n^*.$$

- $\mathcal{F}_X \succeq_{\text{nSD}_5} \mathcal{F}_Y$ implies:

$$\overline{F}_X \succeq_{\text{nSD}_5} \overline{F}_Y \text{ and } \underline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u) \text{ for every } u \in \mathcal{U}_n^*.$$

- $\mathcal{F}_X \succeq_{\text{nSD}_6} \mathcal{F}_Y$ implies:

$$\overline{F}_X \succeq_{\text{nSD}_6} \overline{F}_Y \text{ and } \underline{E}_{\mathcal{F}_X}(u) \geq \underline{E}_{\mathcal{F}_Y}(u) \text{ for every } u \in \mathcal{U}_n^*.$$

Furthermore, the converse implications hold when $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$.

We omit the proof because it is analogous to the one of Proposition 4.19, Theorem 4.23 and Corollaries 4.22 and 4.26.

In the remainder of the subsection we shall investigate several properties of imprecise stochastic dominance. However, from now on we shall focus on the first degree stochastic dominance for two main reasons: on the one hand, it is the most common stochastic dominance in the literature and, on the other hand, as we have just seen, the results for first degree can be easily extended for n -th degree stochastic dominance.

Connection with previous approaches

A first approach to the extension of the stochastic dominance towards an imprecise framework was made by Denoeux in [61].

He considered two random variables U and V such that $P(U \leq V) = 1$. They can be equivalently represented as a random interval $[U, V]$, which in turn induces a belief and a plausibility function, as we saw in Definition 2.43:

$$\text{bel}(A) = P([U, V] \subseteq A) \text{ and } \text{pl}(A) = P([U, V] \cap A \neq \emptyset)$$

for every element A in the Borel sigma-algebra $\beta_{\mathbb{R}}$. Thus, for every $x \in \mathbb{R}$:

$$\text{bel}((-\infty, x]) = F_V(x) \text{ and } \text{pl}((-\infty, x]) = F_U(x).$$

The associated set of probability measures \mathcal{P} compatible with bel and pl is given by:

$$\mathcal{P} = \{P \text{ probability} : bel(A) \leq P(A) \leq pl(A) \text{ for every } A \in \beta_{\mathbb{R}}\}.$$

Denoeux considered two random closed intervals $[U, V]$ and $[U', V']$. One possible way of comparing them is to compare their associated sets of probabilities:

$$\begin{aligned} \mathcal{P} &= \{P \text{ probability} : bel(A) \leq P(A) \leq pl(A) \text{ for every } A \in \beta_{\mathbb{R}}\}. \\ \mathcal{P}' &= \{P \text{ probability} : bel'(A) \leq P(A) \leq pl'(A) \text{ for every } A \in \beta_{\mathbb{R}}\}. \end{aligned}$$

Based on the usual ordering between real intervals (see [78]), Denoeux proposed the following notions:

- $\mathcal{P} \gg \mathcal{P}' \Leftrightarrow pl'((x, \infty)) \leq bel((x, \infty))$ for every $x \in \mathbb{R}$.
- $\mathcal{P} \supseteq \mathcal{P}' \Leftrightarrow pl'((x, \infty)) \leq pl((x, \infty))$ for every $x \in \mathbb{R}$.
- $\mathcal{P} \supseteq \mathcal{P}' \Leftrightarrow bel'((x, \infty)) \leq bel((x, \infty))$ for every $x \in \mathbb{R}$.
- $\mathcal{P} \succeq \mathcal{P}' \Leftrightarrow bel'((x, \infty)) \leq pl((x, \infty))$ for every $x \in \mathbb{R}$.

It turns out that the above notions can be characterised in terms of the stochastic dominance between the lower and upper limits of the random intervals:

Proposition 4.28 ([61]) *Let (U, V) and (U', V') be two pairs of random variables satisfying $P(U \leq V) = P(U' \leq V') = 1$, and let \mathcal{P} and \mathcal{P}' their associated sets of probability measures. The following equivalences hold:*

- $\mathcal{P} \gg \mathcal{P}' \Leftrightarrow U \succeq_{\text{FSD}} V'$.
- $\mathcal{P} \supseteq \mathcal{P}' \Leftrightarrow U \succeq_{\text{FSD}} U'$.
- $\mathcal{P} \supseteq \mathcal{P}' \Leftrightarrow V \succeq_{\text{FSD}} V'$.
- $\mathcal{P} \succeq \mathcal{P}' \Leftrightarrow V \succeq_{\text{FSD}} U'$.

Note that the above definitions can be represented in an equivalent way by means of p-boxes: if we consider the set of distribution functions induced by \mathcal{P} , we obtain

$$\{F : F_V \leq F \leq F_U\},$$

i.e., the p-box determined by F_V and F_U . Similarly, the set \mathcal{P}' induces the p-box $(F_{V'}, F_{U'})$, and Denoeux's definitions are equivalent to comparing the lower and upper distribution functions of these p-boxes, as we can see from Proposition 4.28. Note moreover that the same result holds if we consider finitely additive probability measures

instead of σ -additive ones, because both of them determine the same p-box and the lower and upper distribution functions are included in both cases.

There is a clear connection between the scenario proposed by Denoeux and our proposal. Let $[U, V]$ and $[U', V']$ be two random closed intervals, whose associated belief and plausibility functions determine the sets of probability measures $\mathcal{P}, \mathcal{P}'$ and the sets of cumulative distribution functions \mathcal{F} and \mathcal{F}' . Applying Proposition 4.28 and Corollary 4.22, we obtain the following equivalences:

- $\mathcal{F} \succeq_{\text{FSD}_1} \mathcal{F}' \Leftrightarrow F_U(t) \leq F_{V'}(t)$ for every $t \in \mathbb{R} \Leftrightarrow \mathcal{P} \gg \mathcal{P}'$.
- $\mathcal{F} \succeq_{\text{FSD}_2} \mathcal{F}' \Leftrightarrow \mathcal{F} \succeq_{\text{FSD}_3} \mathcal{F}' \Leftrightarrow F_V(t) \leq F_{V'}(t)$ for every $t \in \mathbb{R} \Leftrightarrow \mathcal{P} \geq \mathcal{P}'$.
- $\mathcal{F} \succeq_{\text{FSD}_4} \mathcal{F}' \Leftrightarrow F_V(t) \leq F_{U'}(t)$ for every $t \in \mathbb{R} \Leftrightarrow \mathcal{P} \gtrsim \mathcal{P}'$.
- $\mathcal{F} \succeq_{\text{FSD}_5} \mathcal{F}' \Leftrightarrow \mathcal{F} \succeq_{\text{FSD}_6} \mathcal{F}' \Leftrightarrow F_U(t) \leq F_{U'}(t)$ for every $t \in \mathbb{R} \Leftrightarrow \mathcal{P} \supseteq \mathcal{P}'$.

Hence, condition \geq gives rise to (FSD_2) (when \mathcal{P} has a smallest distribution function) and (FSD_3) (when it does not have it); similarly, condition \supseteq produces (FSD_5) (if \mathcal{P}' has a greatest distribution function) and (FSD_6) (otherwise).

This also shows that our proposal is more general in the sense that it can be applied to arbitrary sets of probability measures, and not only those associated with a random closed interval. On the other hand, our work is more restrictive in the sense that we are assuming that our referential space is $[0,1]$, instead of the real line. As we mentioned at the beginning of the section, our results are immediately extendable to distribution functions taking values in any closed interval $[a, b]$, where $a < b$ are real numbers. The restriction to bounded intervals is made so that the lower envelope of a set of cumulative distribution functions is a finitely additive distribution function, which may not be the case if we consider the whole real line as our referential space. One solution to this problem is to add to our space a smallest and a greatest value $0_\Omega, 1_\Omega$, so that we always have $F(0_\Omega) = 0$ and $F(1_\Omega) = 1$.

Increasing imprecision

Next we study the behaviour of the different notions of stochastic dominance for sets of distributions when we use them to compare two sets of distribution functions, one of which is more imprecise than the other. This may be useful in some situations: for instance, p-boxes can be seen as *confidence bands* [38, 174], which model our imprecise information about a distribution function taking into account a given sample and a fixed confidence level. Then if we apply two different confidence levels to the same data, we obtain two confidence bands, one included in the other, and we may study which of the two is preferred according to the different criteria we have proposed. In this sense, we may also study our preferences between a set of portfolios, that we represent by means

of a set of distribution functions, and a greater set, where we include more distribution functions, but where also the associated risk may increase.

We are going to consider two different situations: the first one is when our information is given by a set of distribution functions. Hence, we consider two sets $\mathcal{F}_X \subseteq \mathcal{F}_Y$ and investigate our preferences between them:

Proposition 4.29 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y such that $\mathcal{F}_X \subseteq \mathcal{F}_Y$. It holds that:*

1. *If \mathcal{F}_X has only one distribution function, then all the possibilities are valid for (FSD_1) . Otherwise, if \mathcal{F}_X is formed by more than one distribution function, \mathcal{F}_X and \mathcal{F}_Y are incomparable with respect to (FSD_1) .*
2. *With respect to $(FSD_2), \dots, (FSD_6)$, the possible scenarios are summarised in the following table:*

	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X \succ_{FSD_1} \mathcal{F}_Y$				•	•
$\mathcal{F}_Y \succ_{FSD_1} \mathcal{F}_X$	•	•			
$\mathcal{F}_X \equiv_{FSD_1} \mathcal{F}_Y$	•	•	•	•	•
$\mathcal{F}_X, \mathcal{F}_Y$ incomparable	•			•	

Proof: Let us prove that the possibilities ruled out in the statement of the proposition cannot happen:

1. On the one hand, if \mathcal{F}_X has more than one cumulative distribution function, we deduce that \mathcal{F}_X is incomparable with itself with respect to (FSD_1) , and as a consequence it is also incomparable with respect to the greater set \mathcal{F}_Y .
2. Since $\mathcal{F}_X \subseteq \mathcal{F}_Y$, for any $F_1 \in \mathcal{F}_X$ there exists $F_2 \in \mathcal{F}_Y$ such that $F_1 = F_2$. Hence, we always have $\mathcal{F}_Y \succeq_{FSD_3} \mathcal{F}_X$ and $\mathcal{F}_X \succeq_{FSD_6} \mathcal{F}_Y$. Thus, we obtain that $\mathcal{F}_X \not\prec_{FSD_3} \mathcal{F}_Y$, $\mathcal{F}_Y \not\prec_{FSD_6} \mathcal{F}_X$, and both sets cannot be incomparable with respect to (FSD_3) and (FSD_6) . Moreover, using Proposition 4.3 $\mathcal{F}_X \succ_{FSD_2} \mathcal{F}_Y$ and $\mathcal{F}_Y \succ_{FSD_5} \mathcal{F}_X$ are not possible. This also shows that $\mathcal{F}_X \equiv_{FSD_4} \mathcal{F}_Y$, because any $F \in \mathcal{F}_X \subseteq \mathcal{F}_Y$ is equivalent to itself. ■

Next example shows that all the other scenarios are indeed possible.

Example 4.30 • *Let us see that $\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$ is possible for $i = 1, 5, 6$. For this aim, take $\mathcal{F}_X = \{F\}$ and $\mathcal{F}_Y = \{F, F_{1,0}\}$. Then, it holds that $\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$ for $i = 1, 5, 6$ and $\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$ for $i = 2, 3$.*

- Let us check that $\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$, is possible for $i = 1, 2, 3$. Consider $\mathcal{F}_X = \{F\}$ and $\mathcal{F}_Y = \{F, F_{1,1}\}$. Then, it holds that $\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$ for $i = 1, 2, 3$ and $\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$ for $i = 5, 6$.
- Now, let us see that $\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$, is possible for $i = 1, \dots, 6$. For this aim, take $\mathcal{F}_X = \mathcal{F}_Y = \{F\}$. Then, $\mathcal{F}_X \equiv_{FSD_1} \mathcal{F}_Y$ and by Proposition 4.3, $\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$ for any $i = 2, \dots, 6$.
- To see that incomparability is possible for $i = 1, 2, 5$, let $\mathcal{F}_X = \mathcal{F}_Y = \{F, F_{1,0.5}\}$. Then \mathcal{F}_X and \mathcal{F}_Y are (FSD_i) incomparable for $i = 1, 2, 5$, since F and $F_{1,0.5}$ are incomparable. ♦

Remark 4.31 A particular case of the above result would be when we compare a set of distribution functions \mathcal{F}_X with itself, i.e., when $\mathcal{F}_Y = \mathcal{F}_X$. In that case, $\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_X$ for $i = 3, 4, 6$, as we have seen in Proposition 4.9. With respect to (FSD_1) , (FSD_2) and (FSD_5) , we may have either incomparability or indifference: to see that we may have incomparability, consider $\mathcal{F}_X = \mathcal{F}_Y = \{F, F_{1,0.5}\}$; for indifference take $\mathcal{F}_X = \mathcal{F}_Y = \{F\}$. ♦

The second scenario corresponds to the case where our information about the set of distribution functions is given by means of a p-box. A more imprecise p-box corresponds to the case where either the lower distribution function is smaller, the upper distribution function is greater, or both. We begin by considering the latter case.

Proposition 4.32 Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y , and let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ denote their associated p-boxes. Assume that $\underline{F}_Y < \underline{F}_X < \overline{F}_X < \overline{F}_Y$. Then the possible scenarios of stochastic dominance are summarised in the following table:

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$				•	•	•
$\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$		•	•	•		
$\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$				•		
$\mathcal{F}_X, \mathcal{F}_Y$ incomparable	•	•	•	•	•	•

Proof: Using Proposition 4.3, we know that $\mathcal{F}_X \succeq_{FSD_1} \mathcal{F}_Y$ if and only if $\overline{F}_X \leq \underline{F}_Y$, which is incompatible with the assumptions. Similarly, we can see that $\mathcal{F}_Y \not\prec_{FSD_1} \mathcal{F}_X$ and as a consequence they are incomparable.

On the other hand, if $\mathcal{F}_X \succeq_{FSD_i} \mathcal{F}_Y$, for $i = 2, 3$, using Proposition 4.19 it holds that $\underline{F}_X \leq \underline{F}_Y$, a contradiction with the hypothesis.

Similarly, if $\mathcal{F}_Y \succeq_{FSD_i} \mathcal{F}_X$, for $i = 5, 6$, we deduce from Proposition 4.19 that $\overline{F}_Y \leq \overline{F}_X$, again a contradiction. ■

Next example shows that the scenarios included in the table are possible.

Example 4.33 • *Let us see that for (FSD_i) , $i = 2, \dots, 6$, \mathcal{F}_X and \mathcal{F}_Y can be incomparable. For this aim we consider $\mathcal{F}_X = \{F, F^*\}$, where $F^* = \max\{F, F_{1,0.7}\}$, and $\mathcal{F}_Y = \{F_{1,0.5}, F_{\{(0.5,0.5), (0,1)\}}\}$. It is easy to check that both sets of cumulative distribution functions are incomparable, since every distribution function on \mathcal{F}_X is incomparable with every distribution function on \mathcal{F}_Y .*

- *Let us now consider*

$$\mathcal{F}_X = \{F, F^*\} \text{ and } \mathcal{F}_Y = \{F_{(0.5,0.5), (0,0.5)}, F_{(0.5,0.5), (0.5,1)}\}.$$

Then $\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$ for $i = 2, 3$ and $\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$ for $i = 5, 6$. As a consequence, both sets are indifferent with respect to Definition (FSD_4) .

- *Finally, it only remains to see that we may have strict preference under Definition (FSD_4) . On the one hand, if we consider the sets*

$$\mathcal{F}_X = \{F, F^*\} \text{ and } \mathcal{F}_Y = \{F_{1,0.5}, F_{(0.5,0.5), (0,1)}, F_{(0.5,0.5), (0,0.5)}\},$$

it holds that $\mathcal{F}_X \succ_{FSD_4} \mathcal{F}_Y$. In the other hand, if we consider

$$\mathcal{F}_Y = \{F_{1,0.5}, F_{(0.5,0.5), (0,1)}, F_{(0.5,0.5), (0.5,1)}\},$$

we obtain that $\mathcal{F}_Y \succ_{FSD_4} \mathcal{F}_X$.♦

Although the inclusion $\mathcal{F}_X \subseteq \mathcal{F}_Y$ implies that $\underline{E}_Y \leq \underline{E}_X \leq \overline{F}_X \leq \overline{F}_Y$, we may have $\underline{E}_Y < \underline{E}_X < \overline{F}_X < \overline{F}_Y$ even if \mathcal{F}_X and \mathcal{F}_Y are disjoint, for instance when these lower and upper distribution functions are σ -additive and we take the sets $\mathcal{F}_X = \{\underline{E}_X, \overline{F}_X\}$ and $\mathcal{F}_Y = \{\underline{E}_Y, \overline{F}_Y\}$. For this reason in Proposition 4.29 we cannot have $\mathcal{F}_X \succ_{FSD_4} \mathcal{F}_Y$ nor $\mathcal{F}_Y \succ_{FSD_4} \mathcal{F}_X$ and under the conditions of Proposition 4.32 we can.

Proposition 4.34 *Under the above conditions, if in addition $\underline{E}_X, \overline{F}_X$ belong to \mathcal{F}_X and $\underline{E}_Y, \overline{F}_Y$ belong to \mathcal{F}_Y , the possible scenarios are:*

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_1 \succ_{FSD_i} \mathcal{F}_2$					•	•
$\mathcal{F}_2 \succ_{FSD_i} \mathcal{F}_1$		•	•			
$\mathcal{F}_1 \equiv_{FSD_i} \mathcal{F}_2$				•		
$\mathcal{F}_1, \mathcal{F}_2$ incomparable	•					

Proof:

- It is obvious that \mathcal{F}_X and \mathcal{F}_Y are incomparable with respect to Definition (FSD_1) .

- It holds that $\underline{F}_Y < \underline{F}_X \leq F_1$ for any $F_1 \in \mathcal{F}_X$, and then $\mathcal{F}_Y \succ_{FSD_2} \mathcal{F}_X$. Moreover, using Corollary 4.22 (FSD_2) and (FSD_3) are equivalent, and consequently $\mathcal{F}_Y \succ_{FSD_3} \mathcal{F}_X$.
- We know that $\underline{F}_Y < \underline{F}_X$, then $\mathcal{F}_Y \succeq_{FSD_4} \mathcal{F}_X$, and moreover $\overline{F}_X < \overline{F}_Y$, and then $\mathcal{F}_X \succeq_{FSD_4} \mathcal{F}_Y$. Using both inequalities we obtain that $\mathcal{F}_X \equiv_{FSD_4} \mathcal{F}_Y$.
- It holds that $F_1 \leq \overline{F}_X < \overline{F}_Y$ for any $F_1 \in \mathcal{F}_X$, and then $\mathcal{F}_X \succ_{FSD_5} \mathcal{F}_Y$. Furthermore, using Corollary 4.22, (FSD_5) and (FSD_6) are equivalent, and consequently $\mathcal{F}_X \succ_{FSD_6} \mathcal{F}_Y$. ■

In particular, the above result is applicable when $\mathcal{F}_X = (\underline{F}_X, \overline{F}_X)$ and $\mathcal{F}_Y = (\underline{F}_Y, \overline{F}_Y)$, with $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$.

To conclude this part, we consider the case where only one of the bounds becomes more imprecise in the second p-box.

Proposition 4.35 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y , and let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ denote their associated p-boxes.*

a) *Let us assume that $\underline{F}_Y < \underline{F}_X < \overline{F}_X = \overline{F}_Y$. Then the possible scenarios are:*

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X \succ_{FSD_1} \mathcal{F}_Y$				•	•	•
$\mathcal{F}_Y \succ_{FSD_1} \mathcal{F}_X$		•	•	•	•	•
$\mathcal{F}_X \equiv_{FSD_1} \mathcal{F}_Y$				•	•	•
$\mathcal{F}_X, \mathcal{F}_Y$ incomparable	•	•	•	•	•	•

b) *Let us assume that $\underline{F}_Y = \underline{F}_X < \overline{F}_X < \overline{F}_Y$. Then the possible situations are:*

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X \succ_{FSD_1} \mathcal{F}_Y$		•	•	•	•	•
$\mathcal{F}_Y \succ_{FSD_1} \mathcal{F}_X$		•	•	•		
$\mathcal{F}_X \equiv_{FSD_1} \mathcal{F}_Y$		•	•	•		
$\mathcal{F}_X, \mathcal{F}_Y$ incomparable	•	•	•	•	•	•

Proof:

- a) Let us first show that incomparability is the only situation possible according to Definition (FSD_1). As proven in Proposition 4.19, $\mathcal{F}_X \succeq_{FSD_1} \mathcal{F}_Y$ if and only if $\overline{F}_X \leq \underline{F}_Y$. But this inequality is not compatible with the hypothesis. For the same reason, the converse inequality, $\overline{F}_Y \leq \underline{F}_X$ is not possible either.

With respect to (FSD_2) , (FSD_3) , note that if $\underline{F}_Y < \underline{F}_X$,

$$\exists x_0 \in [0, 1] \text{ such that } \underline{F}_Y(x_0) = \inf_{F_2 \in \mathcal{F}_Y} F_2(x_0) < \underline{F}_X(x_0)$$

whence there exists $F_2^* \in \mathcal{F}_Y$ such that $F_2^*(x_0) < \underline{F}_X(x_0) \leq F_1(x_0)$ for all $F_1 \in \mathcal{F}_X$. Thus, $F_1 \not\leq F_2^*$ for any $F_1 \in \mathcal{F}_X$ and $\mathcal{F}_X \not\leq_{FSD_3} \mathcal{F}_Y$. Applying Proposition 4.19, $\mathcal{F}_X \not\leq_{FSD_2} \mathcal{F}_Y$.

b) The proof concerning Definition (FSD_1) is analogous to the one in a).

Concerning (FSD_5) , (FSD_6) , note that since $\overline{F}_X < \overline{F}_Y$,

$$\exists x_0 \in [0, 1] \text{ such that } \overline{F}_Y(x_0) = \sup_{F_2 \in \mathcal{F}_Y} F_2(x_0) > \overline{F}_X(x_0),$$

whence there is $F_2^* \in \mathcal{F}_Y$ such that $F_2^*(x_0) > \overline{F}_X(x_0) \geq F_1(x_0)$ for all $F_1 \in \mathcal{F}_X$, then $F_1 \not\geq F_2^*$ for any $F_1 \in \mathcal{F}_X$ and $\mathcal{F}_Y \not\geq_{FSD_6} \mathcal{F}_X$. It also follows from Proposition 4.19 that $\mathcal{F}_Y \not\geq_{FSD_5} \mathcal{F}_X$. ■

Next we give examples showing that when the lower distribution function is smaller in the second p-box and the upper distribution functions coincide, all the possibilities not ruled out in the first table of the previous proposition can arise. Similar examples can be constructed for the case where $\underline{F}_X = \underline{F}_Y$ and $\overline{F}_X < \overline{F}_Y$.

Example 4.36 • We begin by showing that \mathcal{F}_X and \mathcal{F}_Y can be incomparable under any definition (FSD_i) for $i = 2, \dots, 6$. Let us consider the sets:

$$\mathcal{F}_X = \{F_{(0.5-\frac{1}{n}, 0.5, \frac{1}{n}), (0, 0.5, 1)} \mid n \geq 3\} \text{ and } \mathcal{F}_Y = \{F_{1, 0.5}, F_{(0.5, 0.5), (0, 1)}\}.$$

For all $F_1 \in \mathcal{F}_X$ and $F_2 \in \mathcal{F}_Y$ it holds that $F_2 \not\leq_{FSD} F_1$ and $F_1 \not\leq_{FSD} F_2$. Then, \mathcal{F}_X and \mathcal{F}_Y are incomparable according to (FSD_4) , and therefore also according to (FSD_i) for $i = 2, 3, 5, 6$.

- To see that $\mathcal{F}_X, \mathcal{F}_Y$ can be indifferent according to (FSD_4) , (FSD_5) or (FSD_6) , take:

$$\mathcal{F}_X = \{F_{1, 0.5}, F_{(0.5, 0.5), (0, 0.5)}\} \text{ and } \mathcal{F}_Y = \{F_{(0.5, 0.5), (0, 0.5)}, F_{1, 1}\}.$$

Since $\overline{F}_X = \overline{F}_Y = F_{(0.5, 0.5), (0, 0.5)}$ belongs to both sets, they verify that $\mathcal{F}_X \succeq_{FSD_5} \mathcal{F}_Y$ and also $\mathcal{F}_Y \succeq_{FSD_5} \mathcal{F}_X$. Therefore, $\mathcal{F}_X \equiv_{FSD_5} \mathcal{F}_Y$. As a consequence, they are also indifferent according to (FSD_6) and (FSD_4) .

- Next we show that it is also possible that $\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$ for $i = 5, 6$. Let us consider

$$\mathcal{F}_X = \{F_{(1-\frac{1}{n}, \frac{1}{n}), (0, 1)} : n \geq 3\} \text{ and } \mathcal{F}_Y = \{F_{1, 0}, F_{1, 1}\}.$$

They verify that $\mathcal{F}_X \succeq_{FSD_5} \mathcal{F}_Y$ since $F_{(1-\frac{1}{n}, \frac{1}{n}), (0, 1)} \succeq_{FSD} F_{1, 0}$ for all n ; but $\mathcal{F}_Y \not\leq_{FSD_5} \mathcal{F}_X$ since there is not $F' \in \mathcal{F}_X$ such that $F_{1, 0} \succeq_{FSD} F'$. We conclude that $\mathcal{F}_X \succ_{FSD_5} \mathcal{F}_Y$, and applying Proposition 4.15 also $\mathcal{F}_X \succ_{FSD_6} \mathcal{F}_Y$.

- To see that we may also have $\mathcal{F}_Y \succ_{\text{FSD}_i} \mathcal{F}_X$ for $i = 5, 6$, take:

$$\mathcal{F}_X = \{F_{1,0}, F_{(0.75,0.25),(0,1)}\} \text{ and } \mathcal{F}_Y = \{F_{(1-\frac{1}{n},\frac{1}{n}),(0,1)} : n \geq 3\}.$$

Then $\mathcal{F}_Y \succeq_{\text{FSD}_5} \mathcal{F}_X$ because $F_{(1-\frac{1}{n},\frac{1}{n}),(0,1)} \succeq_{\text{FSD}} F_{1,0}$ for every n , but they are not indifferent with respect to (FSD_5) . Hence, $\mathcal{F}_Y \succ_{\text{FSD}_5} \mathcal{F}_X$ and applying Proposition 4.15 also $\mathcal{F}_Y \succ_{\text{FSD}_6} \mathcal{F}_X$.

- Let us give next an example where $\mathcal{F}_X \succ_{\text{FSD}_4} \mathcal{F}_Y$. Consider

$$\begin{aligned} \mathcal{F}_X &= \{F_{(0.6,0.4),(0.5,1)}, F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)} : n \geq 3\} \text{ and} \\ \mathcal{F}_Y &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}. \end{aligned}$$

Then, $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y$ since $F_{(0.6,0.4),(0.5,1)} \succeq_{\text{FSD}} F_{1,0.5}$ but $\mathcal{F}_Y \not\prec_{\text{FSD}_4} \mathcal{F}_X$ since

$$F_{1,0.5}(0.5) > F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)}(0.5) \text{ for all } n \geq 3$$

and $F_{1,0.5}(0.5) > F_{(0.6,0.4),(0.5,1)}(0.5)$. Also

$$F_{(0.5,0.5),(0,1)}(0) > F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)}(0) \text{ for all } n \geq 3$$

and $F_{(0.5,0.5),(0,1)}(0) > F_{(0.6,0.4),(0.5,1)}(0)$.

- We conclude by showing that it may also happen that $\mathcal{F}_Y \succ_{\text{FSD}_i} \mathcal{F}_X$ for $i = 2, 3, 4$. Let us consider

$$\begin{aligned} \mathcal{F}_X &= \{F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)} : n \geq 3\} \text{ and} \\ \mathcal{F}_Y &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}, F_{(0.5,0.5),(0.5,1)}\}. \end{aligned}$$

It holds that

$$F_{(0.5,0.5),(0.5,1)} \succeq_{\text{FSD}} F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)} \text{ for all } n \geq 3,$$

whence $\mathcal{F}_Y \succeq_{\text{FSD}_i} \mathcal{F}_X$ for $i = 2, 3, 4$. On the other hand,

$$F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)}(0) > F_{(0.5,0.5),(0.5,1)}(0)$$

and

$$F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}),(0,0.5,1)}(0.5) > F_{(0.5,0.5),(0.5,1)}(0.5),$$

whence $\mathcal{F}_X \not\prec_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 2, 3, 4$. ♦

Sets of distribution functions associated with the same p-box

Next we investigate the relationships between the preferences on two sets of distributions functions associated with the same p-box. We consider the case of non-trivial p-boxes (that is, those where the lower and the upper distribution functions are different), since otherwise we obviously obtain indifference.

Proposition 4.37 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y such that $\underline{F}_X = \underline{F}_Y$, $\overline{F}_X = \overline{F}_Y$ and $\underline{F}_X < \overline{F}_X$. Then:*

1. \mathcal{F}_X and \mathcal{F}_Y are incomparable with respect to FSD_1 .
2. With respect to (FSD_i) , $i = 2, \dots, 6$, we may have incomparability, strict stochastic dominance or indifference between \mathcal{F}_X and \mathcal{F}_Y .

Proof: By Proposition 4.19, $\mathcal{F}_X \succeq_{FSD_1} \mathcal{F}_Y$ if and only if $\overline{F}_X \leq \underline{F}_Y$, which in this case holds if and only if $\overline{F}_X = \underline{F}_X$, a contradiction with our hypotheses. ■

With respect to conditions $(FSD_2), \dots, (FSD_6)$, it is easy to find examples of indifference by taking $\mathcal{F}_X = \mathcal{F}_Y$ including the lower and upper distribution functions. Next example shows that we may also have strict dominance or incomparability.

Example 4.38 *In these examples we are going to show that, given two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y associated with the same p -box, then there can be strict dominance or incomparability (that they may also be indifferent has already been showed in Proposition 4.37).*

- *Let us consider*

$$\mathcal{F}_X = \{F_{(0.5,0.5),(0,0.5)}, F_{(0.5,0.5),(0.5,1)}\} \text{ and } \mathcal{F}_Y = \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}.$$

Then, it holds that $\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$ for $i = 2, 3$ and $\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$ for $i = 5, 6$. By reversing the roles of \mathcal{F}_X and \mathcal{F}_Y , we obtain an example of \mathcal{F}_X and \mathcal{F}_Y inducing the same p -box and with $\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$ for $i = 5, 6$ and $\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$ for $i = 2, 3$.

- *To see the incomparability, take*

$$\begin{aligned} \mathcal{F}_X &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\} \text{ and} \\ \mathcal{F}_Y &= \{F_{(\frac{1}{n},0.5,0.5-\frac{1}{n}), (0,0.5,1)}, F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}), (0,0.5,1)} : n \geq 3\}. \end{aligned}$$

It is easy to check that both sets are incomparable with respect to (FSD_4) , and then they are also incomparable with respect to (FSD_i) for $i = 1, \dots, 6$.

- *Finally, if we consider $\mathcal{F}_X = \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}$ and*

$$\mathcal{F}_Y = \{F_{(\frac{1}{n},0.5,0.5-\frac{1}{n}), (0,0.5,1)}, F_{(0.5-\frac{1}{n},0.5,\frac{1}{n}), (0,0.5,1)} : n \geq 3, F_{(0.5,0.5),(0.5,1)}\}.$$

We have that $F_{(0.5,0.5),(0.5,1)} \succeq_{FSD} F_{1,0.5}$, while none of the distribution functions in \mathcal{F}_X is dominated by a distribution function in \mathcal{F}_Y . Thus, $\mathcal{F}_Y \succ_{FSD_4} \mathcal{F}_X$. Again, reversing the roles of \mathcal{F}_X and \mathcal{F}_Y we see that we can also have $\mathcal{F}_Y \succ_{FSD_4} \mathcal{F}_X$. ♦

When the lower and upper distribution functions belong to our set of distributions, we deduce the following result.

Corollary 4.39 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y such that $\underline{F}_X = \underline{F}_Y$, $\overline{F}_X = \overline{F}_Y$, $\underline{F}_X < \overline{F}_X$ and $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X \cap \mathcal{F}_Y$. Then $\mathcal{F}_X \equiv_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 2, \dots, 6$, and they are incomparable with respect to (FSD_1) .*

Proof: The result follows immediately from Proposition 4.37 and Corollary 4.22. \blacksquare

Next we investigate the case where we compare these two sets of distribution functions with a third one, and determine if they produce the same preferences:

Proposition 4.40 *Let us consider \mathcal{F}_X , \mathcal{F}_X^* and \mathcal{F}_Y three sets of cumulative distribution functions such that $\underline{F}_X = \underline{F}_X^*$ and $\overline{F}_X^* = \overline{F}_X$. In that case:*

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X^* \succeq_{\text{FSD}_1} \mathcal{F}_Y$, and $\mathcal{F}_Y \succeq_{\text{FSD}_1} \mathcal{F}_X \Leftrightarrow \mathcal{F}_Y \succeq_{\text{FSD}_1} \mathcal{F}_X^*$.
2. *With respect to definitions $(\text{FSD}_2), \dots, (\text{FSD}_6)$, if we assume that $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$, then the possible scenarios for the relationship between \mathcal{F}_X^* and \mathcal{F}_Y are summarised by the following table:*

	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X^* \succ_{\text{FSD}_i} \mathcal{F}_Y$	•	•	•	•	•
$\mathcal{F}_Y \succ_{\text{FSD}_i} \mathcal{F}_X^*$		•	•		•
$\mathcal{F}_X^* \equiv_{\text{FSD}_i} \mathcal{F}_Y$		•	•	•	•
$\mathcal{F}_X^*, \mathcal{F}_Y$ incomparable	•	•	•		•

Proof: Concerning definition (FSD_1) , Proposition 4.19 assures that $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y$ if and only if $\overline{F}_X^* = \overline{F}_X \leq \underline{F}_Y$, and using the same result this is equivalent to $\mathcal{F}_X^* \succeq_{\text{FSD}_1} \mathcal{F}_Y$. The same result shows that $\mathcal{F}_Y \succeq_{\text{FSD}_1} \mathcal{F}_X$ if and only if $\overline{F}_Y \leq \underline{F}_X = \underline{F}_X^*$, and this is again equivalent to $\mathcal{F}_Y \succeq_{\text{FSD}_1} \mathcal{F}_X^*$.

Let us prove that $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$ and $\mathcal{F}_Y \succeq_{\text{FSD}_2} \mathcal{F}_X^*$ are incompatible. If $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$, then $\mathcal{F}_Y \not\prec_{\text{FSD}_2} \mathcal{F}_X$. This means that for every $F_2 \in \mathcal{F}_Y$ there exist $F_1 \in \mathcal{F}_X$ and x_0 such that $F_1(x_0) < F_2(x_0)$. As a consequence,

$$\inf_{F_1^* \in \mathcal{F}_X^*} F_1^*(x_0) = \underline{F}_X^*(x_0) = \underline{F}_X(x_0) \leq F_1(x_0) < F_2(x_0),$$

whence for every $F_2 \in \mathcal{F}_Y$ there is some $F_1^* \in \mathcal{F}_X^*$ such that $F_1^*(x_0) < F_2(x_0)$, and consequently $F_2 \not\prec F_1^*$. This means that $\mathcal{F}_Y \not\prec_{\text{FSD}_2} \mathcal{F}_X^*$, and therefore we cannot have $\mathcal{F}_Y \succeq_{\text{FSD}_2} \mathcal{F}_X^*$.

Let us show next that $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$ implies that $\mathcal{F}_X^* \succeq_{\text{FSD}_5} \mathcal{F}_Y$. If $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$, there is $F_2 \in \mathcal{F}_Y$ such that $F_1 \leq F_2$ for every $F_1 \in \mathcal{F}_X$. Whence, $\overline{F}_X \leq F_2$, and therefore $\overline{F}_X^* \leq F_2$, which implies that also $F_1^* \leq F_2$ for every $F_1^* \in \mathcal{F}_X^*$. Hence, $\mathcal{F}_X^* \succeq_{\text{FSD}_5} \mathcal{F}_Y$. \blacksquare

Next example shows that the other scenarios are possible.

Example 4.41 Let us consider sets of cumulative distribution functions $\mathcal{F}_X, \mathcal{F}_X^*$ and \mathcal{F}_Y that satisfies $\underline{F}_X = \underline{F}_X^*$ and $\overline{F}_X = \overline{F}_X^*$, and we are going to see that the scenarios given in Proposition 4.40 are possible.

- It is obvious that we can find some examples where $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 2, \dots, 6$ and $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$. To see it, it is enough to consider $\mathcal{F}_X = \mathcal{F}_X^*$.
- Let us show that $\mathcal{F}_X \succ_{\text{FSD}_3} \mathcal{F}_Y$ and $\mathcal{F}_Y \succ_{\text{FSD}_3} \mathcal{F}_X^*$ can hold simultaneously. Consider the sets:

$$\begin{aligned}\mathcal{F}_X &= \{F_{(0.5,0.5),(0.5,1)}, F_{(0.5,0.5),(0.5,1)}\}, \\ \mathcal{F}_X^* &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}, \\ \mathcal{F}_Y &= \{F_{(0.75,0.25),(0.5,1)}, F_{(0.25,0.25,0.5),(0,0.5,1)}\}.\end{aligned}$$

It holds that $\underline{F}_X = \underline{F}_X^*$ and $\overline{F}_X = \overline{F}_X^*$. Moreover it holds that $\mathcal{F}_X \succ_{\text{FSD}_3} \mathcal{F}_Y$ since

$$F_{(0.5,0.5),(0.5,1)} \succeq_{\text{FSD}} F_{(0.75,0.25),(0.5,1)}, F_{(0.25,0.25,0.5),(0,0.5,1)},$$

but for $F_{(0.5,0.5),(0.5,1)}$ there is no distribution function in \mathcal{F}_Y smaller than or equal to $F_{(0.5,0.5),(0.5,1)}$. Similarly, $\mathcal{F}_Y \succ_{\text{FSD}_3} \mathcal{F}_X^*$, since

$$\begin{aligned}F_{(0.75,0.25),(0.5,1)} &\succeq_{\text{FSD}} F_{1,0.5} \text{ and} \\ F_{(0.25,0.25,0.5),(0,0.5,1)} &\succeq_{\text{FSD}} F_{(0.5,0.5),(0,1)}.\end{aligned}$$

However, $F_{1,0.5}, F_{(0.5,0.5),(0,1)} \not\succeq_{\text{FSD}} F_{(0.25,0.25,0.5),(0,0.5,1)}$.

- We now prove that the same can happen with Definition (FSD₆). Let us consider

$$\mathcal{F}_Y = \{F_{(0.25,0.75),(0,0.5)}, F_{(0.5,0.25,0.25),(0,0.5,1)}\}.$$

Then it holds that $\mathcal{F}_X^* \succ_{\text{FSD}_6} \mathcal{F}_Y$ and $\mathcal{F}_Y \succ_{\text{FSD}_6} \mathcal{F}_X$. To check that $\mathcal{F}_X^* \succ_{\text{FSD}_6} \mathcal{F}_Y$ it suffices to see that:

$$\begin{aligned}F_{1,0.5} &\succeq_{\text{FSD}} F_{(0.25,0.75),(0,0.5)} \text{ and that} \\ F_{(0.5,0.5),(0,1)} &\succeq_{\text{FSD}} F_{(0.5,0.25,0.25),(0,0.5,1)},\end{aligned}$$

but $F_{(0.25,0.75),(0,0.5)} \not\succeq_{\text{FSD}} F_{1,0.5}, F_{(0.5,0.5),(0,1)}$. To check that $\mathcal{F}_Y \succ_{\text{FSD}_6} \mathcal{F}_X$ it suffices to see that

$$F_{(0.25,0.75),(0,0.5)}, F_{(0.5,0.25,0.25),(0,0.5,1)} \succeq_{\text{FSD}} F_{(0.5,0.5),(0,0.5)}$$

but $F_{(0.5,0.5),(0,0.5)}$ is not stochastically dominated by none of the distribution in \mathcal{F}_Y .

- Next we prove that it is possible that $\mathcal{F}_X \succ_{\text{FSD}_4} \mathcal{F}_Y$ and $\mathcal{F}_Y \succ_{\text{FSD}_4} \mathcal{F}_X^*$. For this aim, we consider:

$$\begin{aligned}\mathcal{F}_X &= \{F_{(0.25,0.25,0.5),(0,0.5,1)}, F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}, \\ \mathcal{F}_Y &= \{F_{(0.25,0.5,0.25),(0,0.5,1)}, F_{(0.4,0.2,0.4),(0,0.5,1)}\} \text{ and} \\ \mathcal{F}_X^* &= \{F_{(0.25,0.75),(0,0.5)}, F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}.\end{aligned}$$

It holds that $\underline{F}_X = \underline{F}_X^*$ and $\overline{F}_X = \overline{F}_X^*$. Also

$$F_{(0.25,0.25,0.5),(0,0.5,1)} \succeq_{\text{FSD}} F_{(0.25,0.5,0.25),(0,0.5,1)},$$

but no distribution in \mathcal{F}_Y is dominated by a distribution function in \mathcal{F}_X . Whence $\mathcal{F}_X \succ_{\text{FSD}_4} \mathcal{F}_Y$. On the other hand,

$$\begin{aligned} & F_{(0.25,0.5,0.25),(0,0.5,1)} \succeq_{\text{FSD}} F_{(0.25,0.75),(0,0.5)}, \text{ but} \\ & F_{(0.25,0.75),(0,0.5)} \not\succeq_{\text{FSD}} F_{(0.25,0.5,0.25),(0,0.5,1)}, F_{(0.4,0.2,0.4),(0,0.5,1)} \\ & F_{1,0.5} \not\succeq_{\text{FSD}} F_{(0.25,0.5,0.25),(0,0.5,1)}, F_{(0.4,0.2,0.4),(0,0.5,1)}, \\ & F_{(0.5,0.5),(0,1)} \not\succeq_{\text{FSD}} F_{(0.25,0.5,0.25),(0,0.5,1)}, F_{(0.4,0.2,0.4),(0,0.5,1)}, \end{aligned}$$

so $\mathcal{F}_Y \succ_{\text{FSD}_4} \mathcal{F}_X^*$.

- Let us now show that \mathcal{F}_X may strictly dominate \mathcal{F}_Y while \mathcal{F}_X^* and \mathcal{F}_Y are indifferent when we consider definition (FSD_i) for $i = 3, 4, 6$. For this aim consider $\mathcal{F}_X, \mathcal{F}_Y$ associated with the same p -box and such that $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 3, \dots, 6$, as in Example 4.38, and let $\mathcal{F}_X^* = \mathcal{F}_Y$.
- To see that $\mathcal{F}_X^* \equiv_{\text{FSD}_5} \mathcal{F}_Y$ and $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$, it is enough to consider the sets $\mathcal{F}_X = \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}$, $\mathcal{F}_Y = \{F_{(0.5,0.5),(0,0.5)}, F_{(0.5,0.5),(0.5,1)}\}$ and $\mathcal{F}_X^* = \mathcal{F}_Y$.
- For $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$ while $\mathcal{F}_X^*, \mathcal{F}_Y$ are (FSD_i) incomparable for $i = 2, 3, 4$, take

$$\begin{aligned} \mathcal{F}_X &= \{F_{(0.5,0.5),(0.5,1)}, F_{(0.5,0.5),(0,0.5)}\}, \\ \mathcal{F}_Y &= \{F\}, \text{ and} \\ \mathcal{F}_X^* &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}. \end{aligned}$$

- For $\mathcal{F}_X \succ_{\text{FSD}_6} \mathcal{F}_Y$ while $\mathcal{F}_X^*, \mathcal{F}_Y$ are (FSD_6) incomparable, take

$$\begin{aligned} \mathcal{F}_X &= \{F_{(\frac{1}{n}, 1 - \frac{2}{n}, \frac{1}{n}), (0,0.5,1)}, F_{(\frac{1}{2} - \frac{1}{n}, \frac{2}{n}, \frac{1}{2} - \frac{1}{n}), (0,0.5,1)} \mid n \geq 3\}, \\ \mathcal{F}_X^* &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}, \\ \mathcal{F}_Y &= \{F_{(0.5 - \frac{1}{n}, 0.5, \frac{1}{n}), (0,0.5,1)}, F \mid n \geq 3\}. \blacklozenge \end{aligned}$$

Remark 4.42 Note that, under the conditions of the previous proposition, if we assume in addition that $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X \cap \mathcal{F}_X^*$ and that $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$, then we deduce from Corollary 4.22 that $\mathcal{F}_X \succeq_{\text{FSD}_i} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_X^* \succeq_{\text{FSD}_i} \mathcal{F}_Y$, for $i = 1, \dots, 6$. \blacklozenge

σ -additive VS finitely additive distribution functions

Although in this work we are focusing on sets of distribution functions associated with σ -additive probability measures, it is not uncommon to encounter situations where our imprecise information is given by means of sets of *finitely* additive probabilities: this is the case of the models of coherent lower and upper previsions in [205], and in particular

of almost all models of non-additive measures considered in the literature [126]; in this sense they are easier to handle than sets of σ -additive probability measures, which do not have an easy characterisation in terms of their lower and upper envelopes, as showed in [102].

A finitely additive probability measure induces a finitely additive distribution function, and conversely, any finitely additive distribution function can be induced by a finitely additive probability measure [133]. As a consequence, given a p-box $(\underline{F}, \overline{F})$, the set of finitely additive probabilities compatible with this p-box induces the class of finitely additive distribution functions

$$\mathcal{F}' := \{F \text{ finitely additive distribution function} : \underline{F} \leq F \leq \overline{F}\}. \quad (4.4)$$

In particular, both $\underline{F}, \overline{F}$ belong to \mathcal{F}' . Taking this into account, if we define conditions of stochastic dominance analogous to those in Definition 4.1 for sets of finitely additive distribution functions, it is not difficult to establish a characterisation similar to Corollary 4.22.

Lemma 4.43 *Let $\mathcal{F}'_X, \mathcal{F}'_Y$ be two sets of finitely additive distribution functions with associated p-boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$. Assume $\underline{F}_X, \overline{F}_X \in \mathcal{F}'_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}'_Y$.*

1. $\mathcal{F}'_X \succeq_{\text{FSD}_1} \mathcal{F}'_Y \Leftrightarrow \overline{F}_X \leq \underline{F}_Y$.
2. $\mathcal{F}'_X \succeq_{\text{FSD}_2} \mathcal{F}'_Y \Leftrightarrow \mathcal{F}'_X \succeq_{\text{FSD}_3} \mathcal{F}'_Y \Leftrightarrow \underline{F}_X \leq \underline{F}_Y$.
3. $\mathcal{F}'_X \succeq_{\text{FSD}_4} \mathcal{F}'_Y \Leftrightarrow \underline{F}_X \leq \overline{F}_Y$.
4. $\mathcal{F}'_X \succeq_{\text{FSD}_5} \mathcal{F}'_Y \Leftrightarrow \mathcal{F}'_X \succeq_{\text{FSD}_6} \mathcal{F}'_Y \Leftrightarrow \overline{F}_X \leq \overline{F}_Y$.

Proof: The proof is analogous to the one for Corollary 4.22. ■

We deduce in particular that under the above conditions definitions (FSD_2) and (FSD_3) are equivalent, and the same applies to (FSD_5) and (FSD_6) . Note that, although in this result we are using that the lower and upper distribution functions of the p-box belong to the associated set of finitely additive distribution functions, this is not necessary for the first statement.

In this section, we are going to investigate the relationship between the results we have obtained for sets of σ -additive probability measures and those we would obtain for finitely additive ones. Let $\mathcal{P}_X, \mathcal{P}_Y$ be two sets of σ -additive probability measures, and let $\mathcal{F}_X, \mathcal{F}_Y$ be their associated sets of distribution functions. These sets of distribution functions determine p-boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$. Let $\mathcal{F}'_X, \mathcal{F}'_Y$ be two sets of finitely additive distribution functions associated with the p-boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$.

When the lower and upper distribution functions of the associated p-box belong to our set of cumulative distribution functions, we can easily show that the stochastic

dominance holds under the same conditions regardless of whether we work with finitely or σ -additive probability measures:

Corollary 4.44 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y with associated p -boxes $(\underline{E}_X, \overline{F}_X)$, $(\underline{E}_Y, \overline{F}_Y)$, and let $\mathcal{F}'_X, \mathcal{F}'_Y$ be the sets of finitely additive distribution functions associated with these p -boxes. If $\underline{E}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{E}_Y, \overline{F}_Y \in \mathcal{F}_Y$, it holds that:*

$$\mathcal{F}_X \succeq_{\text{FSD}_i} \mathcal{F}_Y \Leftrightarrow \mathcal{F}'_X \succeq_{\text{FSD}_i} \mathcal{F}'_Y,$$

for $i = 1, \dots, 6$.

Proof: The result is an immediate consequence of Corollary 4.22 and Lemma 4.43. ■

However, when the lower and the upper distribution functions induced by \mathcal{F}_X and \mathcal{F}_Y do not belong to these sets, the equivalence no longer holds. We can nonetheless establish the following result:

Proposition 4.45 *Let us consider two sets of cumulative distribution functions \mathcal{F}_X and \mathcal{F}_Y , and two sets of finite distribution functions \mathcal{F}'_X and \mathcal{F}'_Y such that $\mathcal{F}_X, \mathcal{F}'_X$ induce the same p -box $(\underline{E}_X, \overline{F}_X)$ and $\mathcal{F}_Y, \mathcal{F}'_Y$ induce the same p -box $(\underline{E}_Y, \overline{F}_Y)$. Then:*

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \mathcal{F}'_X \succeq_{\text{FSD}_1} \mathcal{F}'_Y$.
2. The relationship $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$ does not have any implication in general on the relationship between \mathcal{F}'_X and \mathcal{F}'_Y with respect to (FSD_i) , for $i = 2, 3, 4, 5, 6$.

Proof:

1. From Proposition 4.19, we know that $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \overline{F}_X \leq \underline{E}_Y$. The same proof allows to show the equivalence with $\mathcal{F}'_X \succeq_{\text{FSD}_1} \mathcal{F}'_Y$.
2. If we apply Proposition 4.40 with $\mathcal{F}'_Y = \mathcal{F}_Y$, we see that all we need to prove is that $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$ is compatible with $\mathcal{F}'_Y \succ_{\text{FSD}_i} \mathcal{F}'_X$ for $i = 2, 5$, with $\mathcal{F}'_X \equiv_{\text{FSD}_i} \mathcal{F}'_Y$ for $i = 2$ and with $\mathcal{F}'_X, \mathcal{F}'_Y$ incomparable with respect to (FSD_5) . ■

Next we give examples of all the possibilities in the previous result.

Example 4.46 *Let us show that, given two sets $\mathcal{F}_X, \mathcal{F}'_X$, with $(\underline{E}_X, \overline{F}_X) = (\underline{E}'_X, \overline{F}'_X)$, and $\mathcal{F}_Y, \mathcal{F}'_Y$, with $(\underline{E}_Y, \overline{F}_Y) = (\underline{E}'_Y, \overline{F}'_Y)$, a number of preference scenarios are possible (the other possible scenarios have already been established in the proof).*

We begin by showing that we may have $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$ and $\mathcal{F}'_Y \succ_{\text{FSD}_2} \mathcal{F}'_X$. To see this, consider $\mathcal{F}_X, \mathcal{F}_Y$ defined by:

$$\begin{aligned}\mathcal{F}_X &= \{F_{(0.5,0.5),(0,0.5)}, F_{(0.5,0.5),(0.5,1)}\} \text{ and} \\ \mathcal{F}_Y &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}.\end{aligned}$$

They are associated with the same p -box and satisfy $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$. We also consider $\mathcal{F}'_X = \mathcal{F}_Y, \mathcal{F}'_Y = \mathcal{F}_X$. A similar reasoning shows that we may have $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$ while $\mathcal{F}'_Y \succ_{\text{FSD}_5} \mathcal{F}'_X$.

Next, we show that we may have $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$ and $\mathcal{F}'_X \equiv_{\text{FSD}_2} \mathcal{F}'_Y$. Let

$$\begin{aligned}\mathcal{F}_X &= \mathcal{F}'_X = \mathcal{F}'_Y = \{F_{(0.5,0.5),(0,0.5)}, F_{(0.5,0.5),(0.5,1)}\} \text{ and} \\ \mathcal{F}_Y &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}.\end{aligned}$$

It can be easily seen that $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$ and that $\mathcal{F}_X, \mathcal{F}_Y$ induce the same p -box. Since $F_{(0.5,0.5),(0.5,1)} \in \mathcal{F}'_X \cap \mathcal{F}'_Y$ satisfies that $F_{(0.5,0.5),(0.5,1)} \leq F_{(0.5,0.5),(0,0.5)}$, we deduce that $\mathcal{F}'_X \equiv_{\text{FSD}_2} \mathcal{F}'_Y$.

To conclude, we give an example where $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$ while $\mathcal{F}'_X, \mathcal{F}'_Y$ are incomparable with respect to (FSD_5) . Consider the sets cumulative distribution functions

$$\begin{aligned}\mathcal{F}_X &= \mathcal{F}'_X = \{F_{(\frac{1}{n}, 1-\frac{2}{n}, \frac{1}{n}), (0,0.5,1)} \mid n \geq 3\}, \\ \mathcal{F}_Y &= \{F_{(0.5,0.5),(0,0.5)}, F_{(0.5,0.5),(0.5,1)}\} \text{ and} \\ \mathcal{F}'_Y &= \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}.\end{aligned}$$

Then $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$ because $F_{(\frac{1}{n}, 1-\frac{2}{n}, \frac{1}{n}), (0,0.5,1)} \leq F_{(0.5,0.5),(0,0.5)}$ for every $n \geq 3$. On the other hand, \mathcal{F}'_X and \mathcal{F}'_Y are incomparable with respect to (FSD_5) . \blacklozenge

It is known that any finitely additive cumulative distribution function F can be approximated by a σ -additive cumulative distribution function F^* : its right-continuous approximation, given by

$$F^*(x) = \inf_{y>x} F(y) \quad \forall x < 1, \quad F^*(1) = 1. \quad (4.5)$$

Hence, to any set \mathcal{F} of finitely additive cumulative distribution functions we can associate a set \mathcal{F}^* of σ -additive cumulative distribution functions, defined by $\mathcal{F}^* := \{F^* : F \in \mathcal{F}\}$, and where F^* is given by Equation (4.5). However, both sets do not model the same preferences, as we can see from the following result:

Proposition 4.47 *Let \mathcal{F} be a set of finitely additive cumulative distribution functions, and let \mathcal{F}^* be the set of their σ -additive approximations. The relationships between \mathcal{F} and \mathcal{F}^* are summarised in the following table:*

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F} \succ_{FSD_i} \mathcal{F}^*$	•	•	•	•	•	•
$\mathcal{F}^* \succ_{FSD_i} \mathcal{F}$						
$\mathcal{F} \equiv_{FSD_i} \mathcal{F}^*$	•	•	•	•	•	•
$\mathcal{F}, \mathcal{F}^*$ incomparable	•	•			•	

Proof: From Equation (4.5), $F \leq F^*$ for any $F \in \mathcal{F}$, whence $\mathcal{F} \succeq_{FSD_i} \mathcal{F}^*$, for $i = 3, 4, 6$. We deduce from Proposition 4.3 that we cannot have $\mathcal{F}^* \succ_{FSD_i} \mathcal{F}$ for $i = 1, \dots, 6$. ■

Next example shows that the remaining scenarios are possible.

Example 4.48 If F_1 is a σ -additive distribution function and we take $\mathcal{F} = \{F_1\}$, we obtain $\mathcal{F}^* = \mathcal{F} = \{F_1\}$, and $\mathcal{F}^* \equiv_{FSD_i} \mathcal{F}$ for $i = 1, \dots, 6$.

On the other hand, if $F_1 = I_{(0.5,1]}$ and $\mathcal{F} = \{F_1\}$, we obtain that $F_1^* = I_{[0.5,1]}$, whence $F_1 < F_1^*$ and as a consequence $\mathcal{F} \succ_{FSD_i} \mathcal{F}^*$ for $i = 1, \dots, 6$.

Finally, if $\mathcal{F} = \{I_{[x,1]} : x \in (0,1)\}$, we obtain that $\mathcal{F}^* = \mathcal{F}$ and \mathcal{F} is incomparable with itself with respect to conditions (FSD_i) for $i = 1, 2, 5$. ♦

Convergence of p-boxes

It is well-known that a distribution function can be seen as the limit of the empirical distribution function that we derive from a sample, as we increase the sample size. Something similar applies when we consider a set of distribution functions: it was proven in [136] that any p-box on the unit interval is the limit of a sequence of p-boxes $(\underline{F}_n, \overline{F}_n)_n$ that are *discrete*, in the sense that for every n both \underline{F}_n and \overline{F}_n have a finite number of discontinuity points.

If for two given p-boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$ we consider respective approximating sequences $(\underline{F}_{X,n}, \overline{F}_{X,n})_n, (\underline{F}_{Y,n}, \overline{F}_{Y,n})_n$, in the sense that

$$\lim_n \underline{F}_{X,n} = \underline{F}_X, \lim_n \overline{F}_{X,n} = \overline{F}_X, \lim_n \underline{F}_{Y,n} = \underline{F}_Y, \lim_n \overline{F}_{Y,n} = \overline{F}_Y,$$

we wonder if it is possible to say something about the preferences between $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ by comparing for each n the discrete p-boxes $(\underline{F}_{X,n}, \overline{F}_{X,n})$ and $(\underline{F}_{Y,n}, \overline{F}_{Y,n})$. This is what we set out to do in this section. We shall be even more general, by considering sets of distribution functions whose associated p-boxes converge to some limit.

Proposition 4.49 Let $(\mathcal{F}_{X,n})_n, (\mathcal{F}_{Y,n})_n$ be two sequences of sets of distribution functions and let us denote their associated sequences of p-boxes by $(\underline{F}_{X,n}, \overline{F}_{X,n})$ and $(\underline{F}_{Y,n}, \overline{F}_{Y,n})$ for $n \in \mathbb{N}$. Let $\mathcal{F}_X, \mathcal{F}_Y$ be two sets of cumulative distribution functions with associated

p -boxes $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$. Let us assume that:

$$\begin{array}{cc} \overline{F}_{X,n} \xrightarrow{n} \overline{F}_X & \underline{F}_{X,n} \xrightarrow{n} \underline{F}_X \\ \overline{F}_{Y,n} \xrightarrow{n} \overline{F}_Y & \underline{F}_{Y,n} \xrightarrow{n} \underline{F}_Y \end{array}$$

and that $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$. Then, $\mathcal{F}_{X,n} \succeq_{\text{FSD}_i} \mathcal{F}_{Y,n} \forall n$, implies that $\mathcal{F}_X \succeq_{\text{FSD}_i} \mathcal{F}_Y$, for $i = 1, \dots, 6$.

Proof: The result immediately follows from Propositions 4.3 and 4.19 and Corollary 4.22. \blacksquare

It follows from the proof above that the assumption that the upper and lower distribution functions belong to the corresponding sets of distribution is not necessary for the implication with respect to (FSD_1) ; however, it is necessary for the other definitions, as we can see in the next example.

Example 4.50 Let us consider the following sets of cumulative distribution functions:

$$\begin{array}{l} \mathcal{F}_X = \{F_{1,0.5}, F_{(0.5,0.5),(0,1)}\}. \\ \mathcal{F}_{X,n} = \{F_{(0.5,0.5),(0,0.5)}, F_{(0.5,0.5),(0.5,1)}\}. \\ \mathcal{F}_Y = \mathcal{F}_{Y,n} = \{F\}. \end{array}$$

\mathcal{F}_X and \mathcal{F}_Y are incomparable with respect to (FSD_4) , and consequently with respect to (FSD_i) , for $i = 1, \dots, 6$. However, $\mathcal{F}_{X,n} \succeq_{\text{FSD}_i} \mathcal{F}_{Y,n}$ for $i = 2, 3, 4$ and $\mathcal{F}_{Y,n} \succeq_{\text{FSD}_i} \mathcal{F}_{X,n}$ for $i = 4, 5, 6$. \blacklozenge

Stochastic dominance between possibility measures

So far, we have explored the extension of the notion of stochastic dominance towards sets of probability measures, and we have showed that in some cases it is equivalent to compare the p -boxes they determine. In this section, we are going to use stochastic dominance to compare possibility measures associated with *continuous* distribution functions. Recall that, from Definition 2.41, a possibility measure Π is a supremum preserving function $\Pi : \mathcal{P}([0, 1]) \rightarrow [0, 1]$, and it is characterised by its restriction to events π , called possibility distribution. Given two possibility measures Π_1 and Π_2 , we can consider their associated credal sets, given by Equation (2.19):

$$\begin{array}{l} \mathcal{M}(\Pi_1) := \{P \text{ probability} : P(A) \leq \Pi_1(A) \forall A\}, \text{ and} \\ \mathcal{M}(\Pi_2) := \{P \text{ probability} : P(A) \leq \Pi_2(A) \forall A\}. \end{array}$$

From these credal sets, we can also consider their associated sets of distribution functions and their associated p -boxes, given in Equation (2.20) by

$$\begin{array}{l} \overline{F}_1(x) = \sup_{y \leq x} \pi_1(y), \quad \underline{F}_1(x) = 1 - \sup_{y > x} \pi_1(y), \\ \overline{F}_2(x) = \sup_{y \leq x} \pi_2(y), \quad \underline{F}_2(x) = 1 - \sup_{y > x} \pi_2(y). \end{array}$$

When considering possibility measures associated with continuous distribution functions, both the lower and the upper distribution functions belong to the set of distribution functions associated with the possibility measures:

Lemma 4.51 *Let Π be a possibility measure associated with a continuous possibility distribution on $[0, 1]$. Then, there exist probability measures $P_1, P_2 \in \mathcal{M}(\Pi)$ whose associated distribution functions are $F_{P_1} = \bar{F}, F_{P_2} = \underline{F}$.*

Proof: Let us consider the probability space $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$, where $\beta_{[0,1]}$ denotes the Borel σ -field and $\lambda_{[0,1]}$ the Lebesgue measure, and let $\Gamma : [0, 1] \rightarrow \mathcal{P}([0, 1])$ be the random set given by $\Gamma(\alpha) = \{x : \pi(x) \geq \alpha\} = \pi^{-1}([\alpha, 1])$. Then it was proved in [84] that Π is the upper probability of Γ .

Let us consider the mappings $U_1, U_2 : [0, 1] \rightarrow [0, 1]$ given by $U_1(\alpha) = \min \Gamma(\alpha)$, $U_2(\alpha) = \max \Gamma(\alpha)$. Since we are assuming that π is a continuous mapping, the set $\pi^{-1}([\alpha, 1]) = \Gamma(\alpha)$ has a maximum and a minimum value for every $\alpha \in [0, 1]$, so U_1, U_2 are well-defined. It also follows that U_1, U_2 are measurable mappings, and as a consequence the probability measures they induce P_{U_1}, P_{U_2} belong to the set $\mathcal{M}(\Pi)$. Their associated distribution functions are:

$$\begin{aligned} F_{U_1}(x) &= P_{U_1}([0, x]) = \lambda_{[0,1]}(U_1^{-1}([0, x])) = \lambda_{[0,1]}(\{\alpha : \min \Gamma(\alpha) \leq x\}) \\ &= \lambda_{[0,1]}(\{\alpha : \exists y \leq x : \pi(y) \geq \alpha\}) = \lambda_{[0,1]}(\{\alpha : \Pi[0, x] \geq \alpha\}) \\ &= \Pi([0, x]) = \bar{F}(x), \end{aligned}$$

where the fifth equality follows from the continuity of $\lambda_{[0,1]}$, and similarly

$$\begin{aligned} F_{U_2}(x) &= P_{U_2}([0, x]) = \lambda_{[0,1]}(U_2^{-1}([0, x])) = \lambda_{[0,1]}(\{\alpha : \max \Gamma(\alpha) \leq x\}) \\ &= \lambda_{[0,1]}(\{\alpha : \pi(y) < \alpha \ \forall y > x\}) = \lambda_{[0,1]}(\{\alpha : \Pi(x, 1] \leq \alpha\}) \\ &= 1 - \Pi((x, 1]) = \underline{F}(x), \end{aligned}$$

again using the continuity of $\lambda_{[0,1]}$. Hence, \bar{F}, \underline{F} belong to the set of distribution functions induced by $\mathcal{M}(\Pi)$. ■

As a consequence, if we consider two possibility measures Π_1, Π_2 with continuous possibility distributions π_1, π_2 , the lower and upper distribution functions of their respective p-boxes belong to the sets $\mathcal{F}_1, \mathcal{F}_2$. Hence, we can apply Proposition 4.21 and conclude that $\mathcal{F}_1 \succeq_{\text{FSD}_2} \mathcal{F}_2 \Leftrightarrow \mathcal{F}_1 \succeq_{\text{FSD}_3} \mathcal{F}_2$ and $\mathcal{F}_1 \succeq_{\text{FSD}_5} \mathcal{F}_2 \Leftrightarrow \mathcal{F}_1 \succeq_{\text{FSD}_6} \mathcal{F}_2$. Moreover, we can use Corollary 4.22 and conclude that:

$$\begin{aligned} \mathcal{F}_1 \succeq_{\text{FSD}_1} \mathcal{F}_2 &\Leftrightarrow \bar{F}_1 \leq \underline{F}_2 \\ \mathcal{F}_1 \succeq_{\text{FSD}_2} \mathcal{F}_2 &\Leftrightarrow \underline{F}_1 \leq \underline{F}_2 \\ \mathcal{F}_1 \succeq_{\text{FSD}_4} \mathcal{F}_2 &\Leftrightarrow \underline{F}_1 \leq \bar{F}_2 \\ \mathcal{F}_1 \succeq_{\text{FSD}_5} \mathcal{F}_2 &\Leftrightarrow \bar{F}_1 \leq \bar{F}_2. \end{aligned}$$

The following proposition gives a sufficient condition for each of these relationships.

Proposition 4.52 *Let $\mathcal{F}_1, \mathcal{F}_2$ be the sets of distribution functions associated with the possibility measures Π_1, Π_2 .*

1. $\Pi_1 \leq N_2 \Rightarrow \mathcal{F}_1 \succeq_{\text{FSD}_1} \mathcal{F}_2$.
2. $\Pi_2 \leq \Pi_1 \Rightarrow \mathcal{F}_1 \succeq_{\text{FSD}_2} \mathcal{F}_2, \mathcal{F}_1 \succeq_{\text{FSD}_3} \mathcal{F}_2$
3. $\mathcal{M}(\Pi_1) \cap \mathcal{M}(\Pi_2) \neq \emptyset \Rightarrow \mathcal{F}_1 \succeq_{\text{FSD}_4} \mathcal{F}_2$.
4. $N_2 \leq N_1 \Rightarrow \mathcal{F}_1 \succeq_{\text{FSD}_5} \mathcal{F}_2, \mathcal{F}_1 \succeq_{\text{FSD}_6} \mathcal{F}_2$.

Proof:

1. Note that $\bar{F}_1 \leq \underline{F}_2$ if and only if $\sup_{y \leq x} \pi_1(y) \leq 1 - \sup_{y > x} \pi_2(y)$ for every x , or, equivalently, if and only if $\Pi_1([0, x]) \leq 1 - \Pi_2((x, 1]) = N_2([0, x])$ for every x . Then, if $\Pi_1(A) \leq N_2(A)$ for any A , in particular the inequality holds for the sets $[0, x]$, and therefore $\bar{F}_1 \leq \underline{F}_2$.
2. Similarly, $\underline{F}_1 \leq \bar{F}_2$ if and only if $1 - \sup_{y \leq x} \pi_1(y) \leq 1 - \sup_{y > x} \pi_2(y)$ for every x , or, equivalently, if and only if $\Pi_2((x, 1]) \leq \Pi_1((x, 1])$ for every x . Then, if $\Pi_2(A) \leq \Pi_1(A)$ for any A , in particular the inequality holds for the sets $(x, 1]$, and therefore $\underline{F}_1 \leq \bar{F}_2$.
3. For the fourth condition of stochastic dominance, note that $\underline{F}_1 \leq \bar{F}_2$ if and only if $1 - \sup_{y > x} \pi_1(y) \leq \sup_{y \leq x} \pi_2(y)$ for every x , or, equivalently, if and only if $1 \leq \Pi_1((x, 1]) + \Pi_2([0, x])$ for every x . As a consequence, if there is a probability $P \in \mathcal{M}(\Pi_1) \cap \mathcal{M}(\Pi_2)$,

$$1 = P((x, 1]) + P([0, x]) \leq \Pi_1((x, 1]) + \Pi_2([0, x]),$$

whence $\mathcal{F}_1 \succeq_{\text{FSD}_4} \mathcal{F}_2$.

4. Finally, note that $\bar{F}_1 \leq \bar{F}_2$ if and only if $\sup_{y \leq x} \pi_1(y) \leq \sup_{y \leq x} \pi_2(y)$ for every x , or, equivalently, if and only if $\Pi_1([0, x]) \leq \Pi_2([0, x])$ for every x . Hence, if $\Pi_1 \leq \Pi_2$ (or, equivalently, if $N_2 \leq N_1$) we have that $\mathcal{F}_1 \succeq_{\text{FSD}_5} \mathcal{F}_2$ and $\mathcal{F}_1 \succeq_{\text{FSD}_6} \mathcal{F}_2$. ■

However, none of the above conditions is necessary, as we show in the next example.

Example 4.53 1. *First of all, let us see that $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \not\Rightarrow \Pi_X \leq N_Y$. For this aim, let π_X, π_Y be given by*

$$\pi_X(x) = \begin{cases} 0 & \text{if } x \leq 0.5 \\ 2x - 1 & \text{otherwise,} \end{cases} \quad \text{and } \pi_Y(x) = \begin{cases} 1 & \text{if } x \leq 0.5 \\ 2 - 2x & \text{otherwise.} \end{cases}$$

Then for every $x \in [0, 1]$ it holds that $\Pi_X([0, x]) + \Pi_Y((x, 1]) \leq 1$: this holds trivially for $x \leq 0.5$ because in that case $\Pi_X([0, x]) = 0$. For $x > 0.5$, we have that

$$\Pi_X([0, x]) + \Pi_Y((x, 1]) = 2x - 1 + 2 - 2x = 1.$$

Hence, $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y$. However:

$$\Pi_X([0.5, 1]) = 1 > \Pi_Y([0.5, 1]) = 1 - \Pi_Y([0, 0.5]) = 1 - 1 = 0,$$

so the converse of the first implication does not hold.

2. Now, we are going to see that $\mathcal{F}_X \succeq_{\text{FSD}_2, \text{FSD}_3} \mathcal{F}_Y \not\Rightarrow \Pi_Y \leq \Pi_X$. Consider the possibility distributions π_X, π_Y given by

$$\pi_X(x) = x, \quad \pi_Y(x) = 1 \quad \forall x.$$

Then $\Pi_Y((x, 1]) = 1 = \Pi_X((x, 1])$ for all x , whence $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y$. However, $\Pi_X([0, 0.5]) = 0.5 < 1 = \Pi_Y([0, 0.5])$, so $\Pi_Y \not\leq \Pi_X$.

3. Now we are going to see that $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \not\Rightarrow \mathcal{M}(\Pi_X) \cap \mathcal{M}(\Pi_Y) \neq \emptyset$. Let π_X, π_Y be given by

$$\pi_X(x) = \begin{cases} 4x - 3 & \text{if } x \geq 0.75 \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad \pi_Y(x) = \begin{cases} 1 - 4x & \text{if } x \leq 0.25 \\ 0 & \text{otherwise.} \end{cases}$$

Then for every $x \in [0, 1]$ it holds that

$$\Pi_X((x, 1]) + \Pi_Y([0, x]) \geq \Pi_Y([0, x]) = 1,$$

whence $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y$. However, any probability P in $\mathcal{M}(\Pi_X) \cap \mathcal{M}(\Pi_Y)$ should satisfy

$$P([0, 0.5]) \leq \Pi_X([0, 0.5]) = 0, \quad P((0.5, 1]) \leq \Pi_Y((0.5, 1]) = 0.$$

Hence, $\mathcal{M}(\Pi_X) \cap \mathcal{M}(\Pi_Y) = \emptyset$.

4. Finally, we are going to see that $\mathcal{F}_X \succeq_{\text{FSD}_5, \text{FSD}_6} \mathcal{F}_Y \not\Rightarrow \Pi_X \leq \Pi_Y$. Consider the possibility distributions π_X, π_Y given by

$$\pi_X(x) = 1, \quad \pi_Y(x) = 1 - x, \quad \forall x.$$

Then it holds that $\Pi_X([0, x]) \leq \Pi_Y([0, x]) \quad \forall x$, whence $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y$. However, $\Pi_X([0.5, 1]) = 1 > 0.5 = \Pi_Y([0.5, 1])$, so $\Pi_X \not\leq \Pi_Y$. ♦

An open problem from this section would be to apply the notion of stochastic dominance to compare possibility measures whose distributions are not necessarily continuous.

P-boxes where one of the bounds is trivial

To conclude this section we investigate the case of p-boxes where one of the bounds is trivial. These have been related to possibility and maxitive measures in [199], and consequently they are in some sense related to the previous paragraph. We shall show that when the lower distribution function is trivial, then the second and third conditions, which are based on the comparison of this bound, always produce indifference.

Proposition 4.54 *Let us consider the p-boxes $\mathcal{F}_X = (\underline{F}_X, \overline{F}_X)$ and $\mathcal{F}_Y = (\underline{F}_Y, \overline{F}_Y)$. Let us assume that $\underline{F}_X = \underline{F}_Y = I_{\{1\}}$, $\underline{F}_X \neq \overline{F}_X$ and $\underline{F}_Y \neq \overline{F}_Y$. Then the possible relationships between \mathcal{F}_X and \mathcal{F}_Y are:*

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$					•	•
$\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$					•	•
$\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$		•	•	•	•	•
$\mathcal{F}_X, \mathcal{F}_Y$ incomparable	•				•	•

Proof:

- Using Proposition 4.19 we know that $\mathcal{F}_X \succeq_{FSD_1} \mathcal{F}_Y \Leftrightarrow \overline{F}_X \leq \underline{F}_Y$. However, this cannot happen since $\underline{F}_Y = I_{\{1\}}$ and the p-boxes are not trivial. Consequently, both sets are incomparable with respect to (FSD_1) .
- Since $\underline{F}_X = \underline{F}_Y \in \mathcal{F}_X \cap \mathcal{F}_Y$, we deduce from Corollary 4.22 that $\mathcal{F}_X \equiv_{FSD_2} \mathcal{F}_Y$. Applying Proposition 4.3, we deduce that $\mathcal{F}_X \equiv_{FSD_3} \mathcal{F}_Y$ and $\mathcal{F}_X \equiv_{FSD_4} \mathcal{F}_Y$.
- On the other hand, it is easy to see that anything can happen for definitions (FSD_5) and (FSD_6) , since these depend on the upper cumulative distribution functions of the p-boxes. ■

Similarly, when the upper distribution function is trivial, then the fifth and sixth conditions, which are based on the comparison of these bounds, always produce indifference.

Proposition 4.55 *Let us consider the p-boxes $\mathcal{F}_X = (\underline{F}_X, \overline{F}_X)$ and $\mathcal{F}_Y = (\underline{F}_Y, \overline{F}_Y)$. Let us assume that $\overline{F}_X = \overline{F}_Y = 1$, $\underline{F}_X < \overline{F}_X$ and $\underline{F}_Y < \overline{F}_Y$. Then the possible relationships between \mathcal{F}_X and \mathcal{F}_Y are:*

	FSD_1	FSD_2	FSD_3	FSD_4	FSD_5	FSD_6
$\mathcal{F}_X \succ_{FSD_i} \mathcal{F}_Y$		•	•			
$\mathcal{F}_Y \succ_{FSD_i} \mathcal{F}_X$		•	•			
$\mathcal{F}_X \equiv_{FSD_i} \mathcal{F}_Y$		•	•	•	•	•
$\mathcal{F}_X, \mathcal{F}_Y$ incomparable	•	•	•			

Proof: This proof is analogous to the previous one. ■

This case is related to the previous paragraph devoted to possibility measures: when the lower distribution function is trivial, the probability measures determined by the p-box are those dominated by the possibility measure that has \overline{F} as a possibility distribution; however, a similar result does not hold for the case of $(\underline{F}, 1)$ in general, because we need \underline{F} to be right-continuous.

0-1-valued p-boxes

Let us now focus on 0-1-valued p-boxes, by which we mean p-boxes where both the lower and upper cumulative distribution functions $\underline{F}, \overline{F}$ are 0-1-valued. As we shall see, the notions of stochastic dominance will be related to the orderings between the intervals of the real line determined by these 0-1-valued distribution functions. 0-1-valued p-boxes have also been related to possibility measures in [199].

Given a 0-1-valued distribution function F , we denote

$$x_F = \inf\{x \mid F(x) = 1\}.$$

Note that this infimum is a minimum when we consider distribution functions associated with σ -additive probability measures, but not necessarily for those associated with finitely additive probability measures.

Using this notation and Proposition 4.19, we can characterise the comparison of sets of 0-1 valued distribution functions:

Proposition 4.56 *Let \mathcal{F}_X and \mathcal{F}_Y be two sets of cumulative distribution functions, with associated p-boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$.*

a) *If $\underline{F}_X, \overline{F}_X, \underline{F}_Y$ and \overline{F}_Y are 0-1-valued functions, then*

1. $\mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow x_{\overline{F}_X} \geq x_{\underline{F}_Y}$.
2. $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Rightarrow x_{\underline{F}_X} \geq x_{\underline{F}_Y}$.
3. $\mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y \Rightarrow x_{\underline{F}_X} \geq x_{\underline{F}_Y}$.
4. $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \Rightarrow x_{\underline{F}_X} \geq x_{\overline{F}_Y}$.
5. $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Rightarrow x_{\overline{F}_X} \geq x_{\overline{F}_Y}$.
6. $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y \Rightarrow x_{\overline{F}_X} \geq x_{\overline{F}_Y}$.

Moreover, if $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$, the converses also hold.

b) *If in particular \mathcal{F}_X and \mathcal{F}_Y are two sets of 0-1 cumulative distribution functions it also holds that*

2. $x_{\underline{F}_X} > x_{\underline{F}_Y} \Leftrightarrow \mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y$.
3. $x_{\underline{F}_X} > x_{\underline{F}_Y} \Leftrightarrow \mathcal{F}_X \succ_{\text{FSD}_3} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y$.
4. $x_{\underline{F}_X} > x_{\overline{F}_Y} \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y$.
5. $x_{\overline{F}_X} > x_{\overline{F}_Y} \Leftrightarrow \mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y$.
6. $x_{\overline{F}_X} > x_{\overline{F}_Y} \Leftrightarrow \mathcal{F}_X \succ_{\text{FSD}_6} \mathcal{F}_Y \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y$.

Proof: In order to prove the first item of this result it is enough to consider Proposition 4.19, and to note that, if F and G are two 0-1 finitely additive distribution functions then $F \leq G$ implies that $x_F \geq x_G$. In particular, if G is a cumulative distribution function, $F \leq G$ if and only if $x_F \geq x_G$, from which we deduce that $x_{\overline{F}_X} \geq x_{\underline{F}_Y} \Rightarrow \mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y$.

Moreover, if $\underline{F}_X, \overline{F}_X \in \mathcal{F}_X$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_Y$, these are cumulative distribution functions, and we can use that $F \leq G$ if and only if $x_F \geq x_G$. Applying Corollary 4.22 we deduce that in that case the converse implications also hold.

Let us consider the second part. On the one hand, it is obvious that $\mathcal{F}_X \succ_{\text{FSD}_i} \mathcal{F}_Y$ implies $\mathcal{F}_X \succeq_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 2, 3, 5, 6$. Let us check the other implications.

2. If $x_{\underline{F}_X} > x_{\underline{F}_Y}$, $\exists x_0$ such that $x_{\underline{F}_X} > x_0 > x_{\underline{F}_Y}$. Then, since $x_0 > x_{\underline{F}_Y}$, $\underline{F}_Y(x_0) = 1$ and therefore $F_2(x_0) = 1 \forall F_2 \in \mathcal{F}_Y$. Since $x_{\underline{F}_X} > x_0$, $\underline{F}_X(x_0) = 0$ and as we are considering only 0 – 1 valued cumulative distribution functions, there is some $F_1 \in \mathcal{F}_X$ such that $F_1(x_0) = 0$. Thus,

$$\exists F_1 \in \mathcal{F}_X \text{ such that } F_1 \succ_{\text{FSD}} F_2 \quad \forall F_2 \in \mathcal{F}_Y.$$

Then, $\mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y$ and $\mathcal{F}_Y \not\prec_{\text{FSD}_2} \mathcal{F}_X$. On the other hand, if $\mathcal{F}_X \succ_{\text{FSD}_2} \mathcal{F}_Y$, Proposition 4.19 implies that $\underline{F}_X \succeq_{\text{FSD}} \underline{F}_Y$, and moreover the preference must be strict (otherwise both sets would be indifferent). Then, $x_{\underline{F}_X} > x_{\underline{F}_Y}$.

3. On the one hand, the direct implication follows from the previous item and Proposition 4.15. On the other hand, if $\mathcal{F}_X \succ_{\text{FSD}_3} \mathcal{F}_Y$, by Proposition 4.19 we know that $\underline{F}_X \succeq_{\text{FSD}} \underline{F}_Y$, and the preference is in fact strict (otherwise the sets \mathcal{F}_X and \mathcal{F}_Y would be indifferent). Then, following the same steps than in the previous item we conclude that $x_{\underline{F}_X} > x_{\underline{F}_Y}$.

4. If $x_{\underline{F}_X} > x_{\overline{F}_Y}$, $\exists x_0$ such that $x_{\underline{F}_X} > x_0 > x_{\overline{F}_Y}$. Then, $\overline{F}_Y(x_0) = 1$, and since all the cumulative distribution function are 0-1 valued, $\exists F_2 \in \mathcal{F}_Y$ such that $F_2(x_0) = 1$. On the other hand, $\underline{F}_X(x_0) = 0$, and since all the cumulative distribution functions are 0-1 valued, there is some $F_1 \in \mathcal{F}_X$ such that $F_1(x_0) = 0$. Hence, $F_1 \leq F_2$ and therefore $\mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y$.

In this case, the preference may be non-strict. For instance, if $\mathcal{F}_X = \mathcal{F}_Y = \{F_1, F_2\}$ such that $x_{F_1} = 0$ and $x_{F_2} = 1$, then $x_{\underline{F}_X} = 1 > 0 = x_{\overline{F}_Y}$ but $\mathcal{F}_X \equiv_{\text{FSD}_4} \mathcal{F}_Y$.

5. If $x_{\overline{F}_X} > x_{\overline{F}_Y}$, there is some x_0 such that $x_{\overline{F}_X} > x_0 > x_{\overline{F}_Y}$. Hence, $\overline{F}_Y(x_0) = 1$. Since all the cumulative distribution functions are 0–1 valued, $\exists F_2 \in \mathcal{F}_Y$ such that $F_2(x_0) = 1$. On the other hand, $\overline{F}_X(x_0) = 0$, whence $F_1(x_0) = 0$ for all $F_1 \in \mathcal{F}_X$. Hence, $F_1 \succ_{\text{FSD}} F_2$ for all $F_1 \in \mathcal{F}_X$. We conclude that $\mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y$ but $\mathcal{F}_Y \not\prec_{\text{FSD}_5} \mathcal{F}_X$.

On the other hand, when $\mathcal{F}_X \succ_{\text{FSD}_5} \mathcal{F}_Y$ Proposition 4.19 implies $\overline{F}_X \succeq_{\text{FSD}} \overline{F}_Y$, and the preference must be strict because otherwise \mathcal{F}_X and \mathcal{F}_Y would be indifferent. Then, $x_{\overline{F}_X} > x_{\overline{F}_Y}$.

6. On the one hand, if $x_{\overline{F}_X} > x_{\overline{F}_Y}$, the result follows from the previous item and Proposition 4.15. On the other hand, when $\mathcal{F}_X \succ_{\text{FSD}_6} \mathcal{F}_Y$, Proposition 4.19 assures that $\overline{F}_X \succeq_{\text{FSD}} \overline{F}_Y$, and the preference must be strict because otherwise \mathcal{F}_X and \mathcal{F}_Y would be indifferent. Then, as we saw in the previous item, it holds that $x_{\overline{F}_X} > x_{\overline{F}_Y}$. ■

Next example shows that the converse implications may not hold in general.

Example 4.57 *We begin by considering the first item. Consider the following sets of distribution functions:*

$$\mathcal{F}_X = \{F_{1,0.5-\frac{1}{n}} : n > 3\} \text{ and } \mathcal{F}_Y = \{F_{1,0.5}\}.$$

It holds that $\underline{F}_X = \underline{F}_Y = \overline{F}_Y = F_{1,0.5}$, and then $x_{\underline{F}_X} = x_{\underline{F}_Y} = 0.5$, but $\mathcal{F}_X \not\prec_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 2, 3, 4$.

Similarly, we can consider the following sets:

$$\mathcal{F}_X = \{F_{1,0.5+\frac{1}{n}} : n > 3\} \text{ and } \mathcal{F}_Y = \{F_{1,0.5}\}.$$

It holds that $\overline{F}_X = \overline{F}_Y = F_{1,0.5}$ and consequently $x_{\overline{F}_X} = x_{\overline{F}_Y} = 0.5$ but $\mathcal{F}_X \not\prec_{\text{FSD}_i} \mathcal{F}_Y$ for $i = 5, 6$.

*We move next to the second item. It is enough to consider a 0-1 valued distribution function F_1 and the sets $\mathcal{F}_X = \mathcal{F}_Y = \{F_1\}$. Both sets are indifferent for Definition (FSD_{*i*}) for $i = 1, \dots, 6$, but no strict inequality hold.◆*

Next we are going to compare the preferences between two sets of 0-1 valued distribution functions and their convex hull. Consider $S_X, S_Y \subseteq [0, 1]$, and let us define the sets:

$$\begin{aligned} \mathcal{F}_{S_X} &= \{F \text{ 0-1 c.d.f.} \mid x_F \in S_X\}. \\ \mathcal{F}_{S_Y} &= \{F \text{ 0-1 c.d.f.} \mid x_F \in S_Y\}. \end{aligned}$$

Since we are working with σ -additive cumulative distribution functions, \mathcal{F}_{S_X} and \mathcal{F}_{S_Y} are related to the degenerate probability measures on elements of S_X, S_Y , respectively.

We shall also consider their convex hulls $\mathcal{F}_X := \text{conv}(\mathcal{F}_{S_X}), \mathcal{F}_Y := \text{conv}(\mathcal{F}_{S_Y})$. These are the sets of cumulative distribution functions with finite supports that are included in S_X and S_Y , respectively.

Now, given any set \mathcal{F} of cumulative distribution functions and its convex hull \mathcal{F}_c , the p-boxes $(\underline{F}, \overline{F})$ and $(\underline{F}_c, \overline{F}_c)$ associated with $\mathcal{F}, \mathcal{F}_c$, coincide:

$$\underline{F} = \underline{F}_c \quad \overline{F} = \overline{F}_c. \quad (4.6)$$

Thus, \mathcal{F}_X and \mathcal{F}_{S_X} determine the same p-box, and the same applies to \mathcal{F}_Y and \mathcal{F}_{S_Y} . We begin with an immediate lemma, whose proof is trivial and therefore omitted.

Lemma 4.58 *Consider $S \subseteq [0, 1]$ and $\mathcal{F}_S = \{F \text{ 0-1 c.d.f.} \mid x_F \in S\}$. Let $\underline{x} = \inf S$ and $\overline{x} = \sup S$ and let $\underline{F}, \overline{F}$ be the lower and upper distribution functions associated with \mathcal{F} . Then*

$$\underline{F} = I_{[\overline{x}, 1]} \quad \text{and} \quad \overline{F} = \begin{cases} I_{[\underline{x}, 1]} & \text{if } \underline{x} \in S, \\ I_{(\underline{x}, 1]} & \text{otherwise.} \end{cases}$$

Moreover, if $\overline{x} \in S$, then $\underline{F} \in \mathcal{F}$, and if $\underline{x} \in S$, then $\overline{F} \in \mathcal{F}$.

Note that when $\overline{F} = I_{(\underline{x}, 1]}$, this is a finite, but not cumulative, distribution function, and as a consequence it cannot belong to \mathcal{F}_S .

Proposition 4.59 *Let S_X and S_Y be two subsets of $[0, 1]$. Then:*

$$1. \mathcal{F}_X \succeq_{\text{FSD}_1} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_1} \mathcal{F}_{S_Y} \Leftrightarrow \inf S_X \geq \sup S_Y.$$

If in addition both $\inf S_X$ and $\sup S_X$ belong to S_X , and also $\inf S_Y$ and $\sup S_Y$ belong to S_Y , then also:

$$2. \mathcal{F}_X \succeq_{\text{FSD}_2} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_2} \mathcal{F}_{S_Y} \Leftrightarrow \max S_X \geq \max S_Y. \text{ Moreover,} \\ \max S_X > \max S_Y \Leftrightarrow \mathcal{F}_{S_X} \succ_{\text{FSD}_2} \mathcal{F}_{S_Y} \text{ and} \\ \max S_X = \max S_Y \Leftrightarrow \mathcal{F}_{S_X} \equiv_{\text{FSD}_2} \mathcal{F}_{S_Y}.$$

$$3. \mathcal{F}_X \succeq_{\text{FSD}_3} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_3} \mathcal{F}_{S_Y} \Leftrightarrow \max S_X \geq \max S_Y. \text{ Moreover,} \\ \max S_X > \max S_Y \Leftrightarrow \mathcal{F}_{S_X} \succ_{\text{FSD}_3} \mathcal{F}_{S_Y} \text{ and} \\ \max S_X = \max S_Y \Leftrightarrow \mathcal{F}_{S_X} \equiv_{\text{FSD}_3} \mathcal{F}_{S_Y}.$$

$$4. \mathcal{F}_X \succeq_{\text{FSD}_4} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_4} \mathcal{F}_{S_Y} \Leftrightarrow \max S_X \geq \min S_Y. \text{ Moreover,} \\ \max S_X > \min S_Y \Leftrightarrow \mathcal{F}_{S_X} \succ_{\text{FSD}_4} \mathcal{F}_{S_Y} \text{ and} \\ \max S_X = \min S_Y \Leftrightarrow \mathcal{F}_{S_X} \equiv_{\text{FSD}_4} \mathcal{F}_{S_Y}.$$

$$5. \mathcal{F}_X \succeq_{\text{FSD}_5} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_5} \mathcal{F}_{S_Y} \Leftrightarrow \min S_X \geq \min S_Y. \text{ Moreover,} \\ \min S_X > \min S_Y \Leftrightarrow \mathcal{F}_{S_X} \succ_{\text{FSD}_5} \mathcal{F}_{S_Y} \text{ and} \\ \min S_X = \min S_Y \Leftrightarrow \mathcal{F}_{S_X} \equiv_{\text{FSD}_5} \mathcal{F}_{S_Y}.$$

6. $\mathcal{F}_X \succeq_{\text{FSD}_6} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_6} \mathcal{F}_{S_Y} \Leftrightarrow \min S_X \geq \min S_Y$. Moreover,
 $\min S_X > \min S_Y \Leftrightarrow \mathcal{F}_{S_X} \succ_{\text{FSD}_6} \mathcal{F}_{S_Y}$ and
 $\min S_X = \min S_Y \Leftrightarrow \mathcal{F}_{S_X} \equiv_{\text{FSD}_6} \mathcal{F}_{S_Y}$.

Proof: The first statement follows from Proposition 4.19 and Equation (4.6), taking also into account that, from Lemma 4.58, $\overline{F}_X \leq \underline{F}_Y$ if and only if $\inf S_X \geq \sup S_Y$.

To prove the other statements, note first of all that if the infima and suprema of S_X and S_Y are included in the set, it follows from Lemma 4.58 that $\underline{F}_X, \overline{F}_X \in \mathcal{F}_{S_X}$ and $\underline{F}_Y, \overline{F}_Y \in \mathcal{F}_{S_Y}$, and applying Corollary 4.22 together with Equation (4.6) we deduce that

$$\mathcal{F}_X \succeq_{\text{FSD}_i} \mathcal{F}_Y \Leftrightarrow \mathcal{F}_{S_X} \succeq_{\text{FSD}_i} \mathcal{F}_{S_Y} \quad \forall i = 2, \dots, 6.$$

On the other hand, it follows from Lemma 4.58 that in those cases

$$\underline{F}_X = I_{[\max S_X, 1]}, \quad \underline{F}_Y = I_{[\max S_Y, 1]}, \quad \overline{F}_X = I_{[\min S_X, 1]}, \quad \overline{F}_Y = I_{[\min S_Y, 1]}.$$

The second and third equivalences in each statement follow then from Corollary 4.22. ■

As a consequence of this result, we obtain the following corollary.

Corollary 4.60 *If S_X and S_Y are closed subsets of $[0, 1]$, then:*

1. $\mathcal{F}_{S_X} \succeq_{\text{FSD}_1} \mathcal{F}_{S_Y} \Leftrightarrow \min S_X \geq \max S_Y$.
2. $\mathcal{F}_{S_X} \succeq_{\text{FSD}_2} \mathcal{F}_{S_Y} \Leftrightarrow \max S_X \geq \max S_Y$.
3. $\mathcal{F}_{S_X} \succeq_{\text{FSD}_3} \mathcal{F}_{S_Y} \Leftrightarrow \max S_X \geq \max S_Y$.
4. $\mathcal{F}_{S_X} \succeq_{\text{FSD}_4} \mathcal{F}_{S_Y} \Leftrightarrow \max S_X \geq \min S_Y$.
5. $\mathcal{F}_{S_X} \succeq_{\text{FSD}_5} \mathcal{F}_{S_Y} \Leftrightarrow \min S_X \geq \min S_Y$.
6. $\mathcal{F}_{S_X} \succeq_{\text{FSD}_6} \mathcal{F}_{S_Y} \Leftrightarrow \min S_X \geq \min S_Y$.

Hence, in that case (FSD_2) is equivalent to (FSD_3) and (FSD_5) is equivalent to (FSD_6) .

It is easy to see that Proposition 4.59 and Corollary 4.60 also hold when we consider \mathcal{F}_X and \mathcal{F}_Y given by

$$\mathcal{F}_X = \{F \text{ c.d.f.} \mid P_F(S_X) = 1\} \text{ and } \mathcal{F}_Y = \{F \text{ c.d.f.} \mid P_F(S_Y) = 1\}.$$

4.1.2 Imprecise statistical preference

In Section 4.1.1 we considered the particular case in which the binary relation is stochastic dominance. Now we focus on the case where the binary relation is that of statistical preference, given in Definition 2.16. Hence, we shall assume that the utility space Ω' is an ordered set, which need not be numerical.

Remark 4.61 *Analogously to the case of stochastic dominance, we shall denote by \succeq_{SP_i} , $i = 1, \dots, 6$ the conditions obtained by using statistical preference as the binary relation in Definition 4.1. We shall also say that \mathcal{X} is (SP_i) preferred or (SP_i) statistically preferred to \mathcal{Y} when $\mathcal{X} \succeq_{SP_i} \mathcal{Y}$. Furthermore, the notation $\mathcal{X} \succeq_{SP_{i,j}} \mathcal{Y}$ means that $\mathcal{X} \succeq_{SP_i} \mathcal{Y}$ and $\mathcal{X} \succeq_{SP_j} \mathcal{Y}$. Note that in Section 4.1.1 we used interchangeably the notation $\mathcal{X} \succeq_{FSD_i} \mathcal{Y}$ and $\mathcal{F}_X \succeq_{FSD_i} \mathcal{F}_Y$, since stochastic dominance is based on the direct comparison of the cumulative distribution functions. Now, we shall only employ the notation $\mathcal{X} \succeq_{SP_i} \mathcal{Y}$, because statistical preference is based on the joint distribution of the random variables, and the marginal distributions do not keep all the information about it. \blacklozenge*

When the binary relation is stochastic dominance, we saw in Proposition 4.15 that there are some general relationships between its strict extensions. In the case of statistical preference, the relationships showed in Proposition 4.15 do not hold in general, as we can see from the following example:

Example 4.62 *Consider the universe $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and let P be the discrete uniform distribution on Ω . Consider the sets of random variables $\mathcal{X} = \{X_1, X_2, X_3\}$ and $\mathcal{Y} = \{X_2, X_4\}$, where the random variables are defined by:*

	ω_1	ω_2	ω_3
X_1	0	2	4
X_2	4	0	2
X_3	2	4	0
X_4	3	2	1

For these sets, since $X_1 \succ_{SP} X_2$ and $X_1 \equiv_{SP} X_4$, then $\mathcal{X} \succeq_{SP_2} \mathcal{Y}$. Moreover, since $X_2 \not\prec_{SP} X_1$ and $X_4 \not\prec_{SP} X_2$, we have that $\mathcal{Y} \not\prec_{SP_2} \mathcal{X}$, hence $\mathcal{X} \succ_{SP_2} \mathcal{Y}$.

However, $\mathcal{X} \not\prec_{SP_3} \mathcal{Y}$: since $X_1 \equiv_{SP} X_4$, $X_2 \equiv_{SP} X_2$ and $X_4 \succ_{SP} X_3$, it holds that $\mathcal{Y} \succeq_{SP_3} \mathcal{X}$. Hence, $\mathcal{X} \equiv_{SP_3} \mathcal{Y}$.

With a similar example it could be proved that $\mathcal{X} \succ_{SP_5} \mathcal{Y}$ and $\mathcal{X} \equiv_{SP_6} \mathcal{Y}$ are compatible statements. \blacklozenge

Note that \succeq_{SP} is reflexive and complete, but it is neither antisymmetric nor transitive. Hence, Proposition 4.6 does not apply in this case; indeed, we can use statistical preference to show that Proposition 4.6 cannot be extended to non transitive relationships.

Example 4.63 Consider the random variables A, B, C from Example 3.83 such that $A \succ_{\text{SP}} B \succ_{\text{SP}} C \succ_{\text{SP}} A$, and let $\mathcal{X} = \{A, B\}, \mathcal{Y} = \{A, C\}$. Then since $A \succeq_{\text{SP}} A$ and $B \succeq_{\text{SP}} C$, we deduce that $\mathcal{X} \succeq_{\text{SP}_3} \mathcal{Y}$; since $A \succ_{\text{SP}} B$ and $C \succ_{\text{SP}} A$, we see that $\mathcal{X} \not\succeq_{\text{SP}_2} \mathcal{Y}$; however, \mathcal{X} has a maximum element, because $A \succ_{\text{SP}} B$. \blacklozenge

On the other hand, since statistical preference complies with Pareto dominance we deduce from Proposition 4.7 that the different conditions can be reduced to the comparison of the maximum and minimum elements of \mathcal{X}, \mathcal{Y} , when these maximum and minimum elements exist. Finally, we deduce from Propositions 4.9 and 4.11 that conditions $\succeq_{\text{SP}_3}, \succeq_{\text{SP}_4}, \succeq_{\text{SP}_6}$ induce a reflexive and complete relationship.

We can also use statistical preference to show that Proposition 4.11 cannot be extended to the relations \succeq_1, \succeq_2 nor \succeq_5 : take the sets $\mathcal{X} = \mathcal{Y} = \{A, B, C\}$, where the variables A, B, C satisfy $A \succ_{\text{SP}} B \succ_{\text{SP}} C \succ_{\text{SP}} A$ as in Example 3.83; then the set \mathcal{X} has neither a maximum nor a minimum element, whence it is incomparable with itself with respect to \succeq_{SP_2} and \succeq_{SP_5} . Applying Proposition 4.3, we deduce that \mathcal{X}, \mathcal{Y} are also incomparable with respect to \succeq_{SP_1} .

We showed in Theorem 4.23 that the generalisations of stochastic dominance towards sets of variables are related to lower and upper expectations. Next, we establish a similar result for the generalisations of statistical preference. Recall that in Theorem 3.40 we proved that:

$$\sup \text{Me}(X - Y) > 0 \Rightarrow X \succeq_{\text{SP}} Y \Rightarrow \sup \text{Me}(X - Y) \geq 0. \quad (4.7)$$

Taking into this result, we shall establish a generalisation in terms of lower and upper medians, and for this we shall require our utility space Ω' to be the reals. Let us consider two sets of alternatives \mathcal{X}, \mathcal{Y} with values on Ω' , and let us introduce the following notation:

$$\begin{aligned} \text{Me}(\mathcal{X} - \mathcal{Y}) &= \{\text{Me}(X - Y) : X \in \mathcal{X}, Y \in \mathcal{Y}\}. \\ \underline{\text{Me}}(\mathcal{X} - \mathcal{Y}) &= \inf \text{Me}(\mathcal{X} - \mathcal{Y}). \\ \overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) &= \sup \text{Me}(\mathcal{X} - \mathcal{Y}), \end{aligned}$$

where we recall that the median of a random variable with respect to a probability measure is given by Equation (3.14).

Proposition 4.64 Let \mathcal{X}, \mathcal{Y} be two sets of random variables defined on a probability space (Ω, \mathcal{A}, P) and taking values on \mathbb{R} .

1. $\underline{\text{Me}}(\mathcal{X} - \mathcal{Y}) > 0 \Rightarrow \mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y} \Rightarrow \overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) \geq 0$.
2. $\exists X \in \mathcal{X}$ such that $\underline{\text{Me}}(\{X\} - \mathcal{Y}) > 0 \Rightarrow \mathcal{X} \succeq_{\text{SP}_2} \mathcal{Y} \Rightarrow \exists X \in \mathcal{X}$ such that $\overline{\text{Me}}(\{X\} - \mathcal{Y}) \geq 0$.
3. $\overline{\text{Me}}(\mathcal{X} - \{Y\}) > 0 \forall Y \in \mathcal{Y} \Rightarrow \mathcal{X} \succeq_{\text{SP}_3} \mathcal{Y} \Rightarrow \overline{\text{Me}}(\mathcal{X} - \{Y\}) \geq 0 \forall Y \in \mathcal{Y}$.

$$4. \overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) > 0 \Rightarrow \mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y} \Rightarrow \overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) \geq 0.$$

$$5. \exists Y \in \mathcal{Y} \text{ such that } \underline{\text{Me}}(\mathcal{X} - \{Y\}) > 0 \Rightarrow \mathcal{X} \succeq_{\text{SP}_5} \mathcal{Y} \Rightarrow \exists Y \in \mathcal{Y} \text{ such that } \overline{\text{Me}}(\mathcal{X} - \{Y\}) \geq 0.$$

$$6. \overline{\text{Me}}(\{X\} - \mathcal{Y}) > 0 \forall X \in \mathcal{X} \Rightarrow \mathcal{X} \succeq_{\text{SP}_6} \mathcal{Y} \Rightarrow \overline{\text{Me}}(\{X\} - \mathcal{Y}) \geq 0 \forall X \in \mathcal{X}.$$

Proof: Recall once more that from Equation (4.7) given two random variables X, Y ,

$$\overline{\text{Me}}(X - Y) > 0 \Rightarrow X \succeq_{\text{SP}} Y \Rightarrow \overline{\text{Me}}(X - Y) \geq 0.$$

SP₁ : If $\underline{\text{Me}}(\mathcal{X} - \mathcal{Y}) > 0$, in particular $\overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) > 0$, and then $\text{Me}(X - Y) > 0$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. Applying Equation (4.7), $X \succeq_{\text{SP}} Y$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, and consequently $\mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y}$. Moreover,

$$\begin{aligned} \mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y} &\Rightarrow X \succeq_{\text{SP}} Y \text{ for every } X \in \mathcal{X}, Y \in \mathcal{Y} \\ &\Rightarrow \sup \text{Me}(X - Y) \geq 0 \text{ for every } X \in \mathcal{X}, Y \in \mathcal{Y} \Rightarrow \overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) \geq 0. \end{aligned}$$

SP₂ : If there is some $X \in \mathcal{X}$ such that $\underline{\text{Me}}(\{X\} - \mathcal{Y}) > 0$, then $\text{Me}(X - Y) > 0$ for every $Y \in \mathcal{Y}$. Applying Equation (4.7), we deduce that $X \succeq_{\text{SP}} Y$ for every $Y \in \mathcal{Y}$, and therefore $\mathcal{X} \succeq_{\text{SP}_2} \mathcal{Y}$.

On the other hand,

$$\begin{aligned} \mathcal{X} \succeq_{\text{SP}_2} \mathcal{Y} &\Rightarrow \text{there is some } X \in \mathcal{X} \text{ such that } X \succeq_{\text{SP}} Y \text{ for every } Y \in \mathcal{Y} \\ &\Rightarrow \sup \text{Me}(X - Y) \geq 0 \text{ for every } Y \in \mathcal{Y} \Rightarrow \overline{\text{Me}}(\{X\} - \mathcal{Y}) \geq 0. \end{aligned}$$

SP₃ : Consider $Y \in \mathcal{Y}$. If $\overline{\text{Me}}(\mathcal{X} - \{Y\}) > 0$, then there is some $X \in \mathcal{X}$ such that $\text{Me}(X - Y) > 0$. Hence, for every $Y \in \mathcal{Y}$ there is $X \in \mathcal{X}$ such that $X \succeq_{\text{SP}} Y$, and consequently $\mathcal{X} \succeq_{\text{SP}_3} \mathcal{Y}$. Moreover,

$$\begin{aligned} \mathcal{X} \succeq_{\text{SP}_3} \mathcal{Y} &\Rightarrow \text{for every } Y \in \mathcal{Y} \text{ there is } X \in \mathcal{X} \text{ such that } X \succeq_{\text{SP}} Y \\ &\Rightarrow \text{for every } Y \in \mathcal{Y} \text{ there is } X \in \mathcal{X} \text{ such that } \sup \text{Me}(X - Y) \geq 0 \\ &\Rightarrow \text{for every } Y \in \mathcal{Y} \text{ it holds that } \overline{\text{Me}}(\mathcal{X} - \{Y\}) \geq 0. \end{aligned}$$

SP₄ : If $\overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) > 0$, there are $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ such that $\text{Me}(X - Y) > 0$, and consequently $X \succeq_{\text{SP}} Y$. Thus, $\mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y}$. On the other hand,

$$\begin{aligned} \mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y} &\Rightarrow \text{there are } X \in \mathcal{X}, Y \in \mathcal{Y} \text{ such that } X \succeq_{\text{SP}} Y \\ &\Rightarrow \text{there are } X \in \mathcal{X}, Y \in \mathcal{Y} \text{ such that } \sup \text{Me}(X - Y) \geq 0 \Rightarrow \overline{\text{Me}}(\mathcal{X} - \mathcal{Y}) \geq 0. \end{aligned}$$

SP₅ : Assume that there exists some $Y \in \mathcal{Y}$ such that $\underline{\text{Me}}(\mathcal{X} - \{Y\}) > 0$. Then $\text{Me}(X - Y) > 0$ for every $X \in \mathcal{X}$, and applying (4.7) we conclude that there is $Y \in \mathcal{Y}$

such that $X \succeq_{\text{SP}} Y$ for every $X \in \mathcal{X}$, and consequently $\mathcal{X} \succeq_{\text{SP}_5} \mathcal{Y}$. On the other hand,

$$\begin{aligned} \mathcal{X} \succeq_{\text{SP}_5} \mathcal{Y} &\Rightarrow \text{there is } Y \in \mathcal{Y} \text{ such that } X \succeq_{\text{SP}} Y \text{ for every } Y \in \mathcal{Y} \\ &\Rightarrow \text{there is } Y \in \mathcal{Y} \text{ such that } \sup \text{Me}(X - Y) \geq 0 \text{ for every } X \in \mathcal{X} \\ &\Rightarrow \text{there is } Y \in \mathcal{Y} \text{ such that } \overline{\text{Me}}(\mathcal{X} - \{Y\}) \geq 0. \end{aligned}$$

SP₆ : Finally, if $\overline{\text{Me}}(\{X\} - \mathcal{Y}) > 0$ for every $X \in \mathcal{X}$, then for every $X \in \mathcal{X}$ there is some $Y \in \mathcal{Y}$ such that $\text{Me}(X - Y) > 0$, whence (4.7) implies that $X \succeq_{\text{SP}} Y$. We conclude that $\mathcal{X} \succeq_{\text{SP}_6} \mathcal{Y}$. Moreover,

$$\begin{aligned} \mathcal{X} \succeq_{\text{SP}_6} \mathcal{Y} &\Rightarrow \text{for every } X \in \mathcal{X} \text{ there is } Y \in \mathcal{Y} \text{ such that } X \succeq_{\text{SP}} Y \\ &\Rightarrow \text{for every } X \in \mathcal{X} \text{ there is } Y \in \mathcal{Y} \text{ such that } \sup \text{Me}(X - Y) \geq 0 \\ &\Rightarrow \text{for every } X \in \mathcal{X}, \overline{\text{Me}}(\{X\} - \mathcal{Y}) \geq 0. \quad \blacksquare \end{aligned}$$

Taking into account the properties of the median, we conclude from this result that statistical preference may be seen as a more robust alternative to stochastic dominance or expected utility in the presence of outliers.

As we made in Section 4.1.1 with imprecise stochastic dominance, now we shall investigate some of the properties of the imprecise statistical preference.

Increasing imprecision

We first study the behavior of conditions \succeq_{SP_i} , $i = 1, \dots, 6$, when we enlarge the sets \mathcal{X}, \mathcal{Y} of alternatives we want to compare. This may correspond to an increase in the imprecision of our models. Not surprisingly, if the more restrictive condition \succeq_{SP_1} is satisfied on the large sets, then it is automatically satisfied on the smaller ones; while for the least restrictive one \succeq_{SP_4} we have the opposite implication.

Proposition 4.65 *Let $\mathcal{X}, \mathcal{Y}, \tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ be four sets of random variables satisfying $\mathcal{X} \subseteq \tilde{\mathcal{X}}$ and $\mathcal{Y} \subseteq \tilde{\mathcal{Y}}$. Then*

$$\tilde{\mathcal{X}} \succeq_{\text{SP}_1} \tilde{\mathcal{Y}} \Rightarrow \mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y} \text{ and } \mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y} \Rightarrow \tilde{\mathcal{X}} \succeq_{\text{SP}_4} \tilde{\mathcal{Y}}.$$

Proof: It is clear that $\tilde{\mathcal{X}} \succeq_{\text{SP}_1} \tilde{\mathcal{Y}} \Rightarrow \mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y}$, since if $X \succeq_{\text{SP}} Y$ for every $X \in \tilde{\mathcal{X}}$ and $Y \in \tilde{\mathcal{Y}}$, the inequality holds in particular for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$.

On the other hand, $\mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y}$ implies the existence of $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ satisfying $X \succeq_{\text{SP}} Y$, and then the inclusions $\mathcal{X} \subseteq \tilde{\mathcal{X}}$ and $\mathcal{Y} \subseteq \tilde{\mathcal{Y}}$ imply that $\tilde{\mathcal{X}} \succeq_{\text{SP}_4} \tilde{\mathcal{Y}}$. \blacksquare

Similar implications cannot be established for \succeq_{SP_i} , for $i = 2, 3, 5, 6$, as the following example shows:

Example 4.66 Consider the universe $\Omega = \{\omega\}$ and let δ_x denote the random variable satisfying $\delta_x(\omega) = x$.

Let us prove that $\mathcal{X} \succ_{\text{SP}_i} \mathcal{Y}$ and $\tilde{\mathcal{Y}} \succ_{\text{SP}_i} \tilde{\mathcal{X}}$ is possible for $i = 2, 3, 5, 6$:

- Consider $\mathcal{X} = \{\delta_0\}$, $\tilde{\mathcal{X}} = \{\delta_0, \delta_2\}$ and $\mathcal{Y} = \tilde{\mathcal{Y}} = \{\delta_1\}$. It holds that $\mathcal{Y} \succ_{\text{SP}_i} \mathcal{X}$ for $i = 1, \dots, 6$ while $\tilde{\mathcal{X}} \succ_{\text{SP}_i} \tilde{\mathcal{Y}}$ for $i = 2, 3$, since $\delta_2 \succeq_{\text{SP}} \delta_1$.
- Now, given $\mathcal{X} = \{\delta_2\}$, $\tilde{\mathcal{X}} = \{\delta_0, \delta_2\}$ and $\mathcal{Y} = \tilde{\mathcal{Y}} = \{\delta_1\}$, it holds that $\mathcal{X} \succ_{\text{SP}_i} \mathcal{Y}$ for $i = 1, \dots, 6$ while $\tilde{\mathcal{Y}} \succ_{\text{SP}_i} \tilde{\mathcal{X}}$ for $i = 5, 6$, since $\delta_1 \succeq_{\text{SP}} \delta_0$.

Note that these examples also show that the implications of the previous proposition are not equivalences in general. \blacklozenge

One particular case when we may enlarge our sets of alternatives is when we consider convex combinations (note that for this we shall again to assume that the utility space Ω' is equal to \mathbb{R}). This may be of interest for instance if we want to compare random sets by means of their measurable selections, as we shall do in Section 4.2.1, and we move from a purely atomic to a non-atomic initial probability space. We shall consider two possibilities, for a given set of alternatives \mathcal{D} : its *convex hull*

$$\widehat{\text{Conv}}(\mathcal{D}) = \left\{ \tilde{U} = \sum_{i=1}^n \lambda_i X_i : \lambda_i > 0, X_i \in \mathcal{D} \forall i, \sum_{i=1}^n \lambda_i = 1 \right\},$$

and also the set of alternatives whose utilities belong to the range of utilities determined by \mathcal{A} :

$$\text{Conv}(\mathcal{D}) = \{ \tilde{U} \text{ r.v.} \mid \tilde{U}(\omega) \in \text{Conv}(\{U(\omega) : U \in \mathcal{D}\}) \}; \quad (4.8)$$

note that $\mathcal{D} \subseteq \widehat{\text{Conv}}(\mathcal{D}) \subseteq \text{Conv}(\mathcal{D})$. Then Proposition 4.65 allows to immediately deduce the following:

Corollary 4.67 Consider two sets of alternatives \mathcal{X}, \mathcal{Y} .

- $\text{Conv}(\mathcal{X}) \succeq_{\text{SP}_1} \text{Conv}(\mathcal{Y}) \Rightarrow \widehat{\text{Conv}}(\mathcal{X}) \succeq_{\text{SP}_1} \widehat{\text{Conv}}(\mathcal{Y}) \Rightarrow \mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y}$.
- $\text{Conv}(\mathcal{X}) \succeq_{\text{SP}_4} \text{Conv}(\mathcal{Y}) \Rightarrow \mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y} \Rightarrow \widehat{\text{Conv}}(\mathcal{X}) \succeq_{\text{SP}_4} \widehat{\text{Conv}}(\mathcal{Y})$.

To see that we cannot establish similar implications with respect to \succeq_{SP_i} , $i = 2, 3, 5, 6$, take the following example:

Example 4.68 Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$ with $P(\{\omega_i\}) = \frac{1}{3}$ for every $i = 1, 2, 3$. Let us consider the sets of variables $\mathcal{X} = \{X_1, X_2\}$ and $\mathcal{Y} = \{Y\}$ given by:

	ω_1	ω_2	ω_3
X_1	0	3	0
X_2	3	0	0
Y	1	1	1

Then since $Q(X_1, Y) = Q(X_2, Y) = \frac{1}{3}$, it follows that $\mathcal{Y} \succ_{\text{SP}_i} \mathcal{X}$ for $i = 1, \dots, 6$. However, $\widehat{\text{Conv}}(\mathcal{X}) \succ_{\text{SP}_i} \widehat{\text{Conv}}(\mathcal{Y})$, for $i = 2, 3$, $\widehat{\text{Conv}}(\mathcal{X}) \equiv_{\text{SP}_4} \widehat{\text{Conv}}(\mathcal{Y})$ and they are incomparable with respect to \succeq_{SP_1} .

On the other hand, if we consider instead the sets $\mathcal{X} = \{X_1, X_2\}$ and $\mathcal{Y} = \{Y\}$, where

	ω_1	ω_2	ω_3
X_1	0	3	3
X_2	3	0	3
Y	2	2	2

it holds that $\mathcal{X} \succ_{\text{SP}_i} \mathcal{Y}$ for $i = 1, \dots, 6$. However, $\widehat{\text{Conv}}(\mathcal{Y}) \succ_{\text{SP}_i} \widehat{\text{Conv}}(\mathcal{X})$, for $i = 5, 6$.

The same sets of variables show that there is no additional implication if we consider the convex hulls determined by Equation (4.8) instead. \blacklozenge

Connection with aggregation functions

Since the binary relation associated with statistical preference is complete, we deduce from Proposition 4.11 that the relations \succeq_{SP_3} , \succeq_{SP_4} , \succeq_{SP_6} also induce a complete relation. Such relations are interesting because they mean that we can always express a preference between two sets of alternatives \mathcal{X}, \mathcal{Y} . One way of deriving a complete relation when we make multiple comparisons is to establish a degree of preference for every pairwise comparison, and to aggregate these degrees of preference into a joint one. This is possible by means of an aggregation function.

Let $\mathcal{X} = \{X_1, \dots, X_n\}$ and $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ be two finite sets of random variables taking values on an ordered utility space Ω' , and let us compute the statistical preference $Q(X_i, Y_j)$ for every pair of variables $X_i \in \mathcal{X}$, $Y_j \in \mathcal{Y}$ by means of Equation (2.7). The set of all these preferences is an instance of *profile of preference* [80], and can be represented by means of the matrix

$$\mathbb{Q}_{\mathcal{X}, \mathcal{Y}} := \begin{pmatrix} Q(X_1, Y_1) & Q(X_1, Y_2) & \dots & Q(X_1, Y_m) \\ \vdots & \vdots & \ddots & \vdots \\ Q(X_n, Y_1) & Q(X_n, Y_2) & \dots & Q(X_n, Y_m) \end{pmatrix} \quad (4.9)$$

Note that the profile of preferences of \mathcal{Y} over \mathcal{X} , $\mathbb{Q}_{\mathcal{Y},\mathcal{X}}$, corresponds to one minus the transposed matrix of $\mathbb{Q}_{\mathcal{X},\mathcal{Y}}$, i.e., $1 - \mathbb{Q}_{\mathcal{X},\mathcal{Y}}^t$. We shall show that conditions $\succeq_{\text{SP}_1}, \dots, \succeq_{\text{SP}_6}$ can be expressed by means of an aggregation function over the profile of preference:

Definition 4.69 ([31, 80]) *An aggregation function is a mapping defined by*

$$G : \cup_{s \in \mathbb{N}} [0, 1]^s \rightarrow [0, 1],$$

that it is componentwise increasing and satisfies the boundary conditions $G(0, \dots, 0) = 0$ and $G(1, \dots, 1) = 1$.

The matrix $\mathbb{Q}_{\mathcal{X},\mathcal{Y}}$ representing the profile of preferences between \mathcal{X} and \mathcal{Y} can be equivalently represented by means of a vector on $[0, 1]^{nm}$ using the lexicographic order:

$$\vec{z}_{\mathcal{X},\mathcal{Y}} = (Q(X_1, Y_1), Q(X_1, Y_2), \dots, Q(X_1, Y_m), Q(X_2, Y_1), \dots, Q(X_n, Y_m)).$$

Taking this into account, given an aggregation function $G : \cup_{s \in \mathbb{N}} [0, 1]^s \rightarrow [0, 1]$, we shall denote by $G(\mathbb{Q}_{\mathcal{X},\mathcal{Y}})$ the image of the vector $\vec{z}_{\mathcal{X},\mathcal{Y}}$ by means of this aggregation function.

Definition 4.70 *Given two finite sets of random variables \mathcal{X} and \mathcal{Y} , $\mathcal{X} = \{X_1, \dots, X_n\}$ and $\mathcal{Y} = \{Y_1, \dots, Y_m\}$, and an aggregation function G , we say that \mathcal{X} is G -statistically preferred to \mathcal{Y} , and denote it by $\mathcal{X} \succeq_{\text{SP}_G} \mathcal{Y}$, if*

$$G(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) := G(\vec{z}_{\mathcal{X},\mathcal{Y}}) \geq \frac{1}{2}. \quad (4.10)$$

We refer to [31] for a review of aggregation functions. Some important properties are the following:

Definition 4.71 ([31]) *An aggregation function $G : \cup_{s \in \mathbb{N}} [0, 1]^s \rightarrow [0, 1]$ is called:*

- Symmetric if it is invariant under permutations.
- Monotone if $G(r_1, \dots, r_s) \geq G(r'_1, \dots, r'_s)$ whenever $r_i \geq r'_i$ for every $i = 1, \dots, s$.
- Idempotent if $G(r, \dots, r) = r$.

We shall call an aggregation function $G : \cup_{s \in \mathbb{N}} [0, 1]^s \rightarrow [0, 1]$ *self-dual* if

$$G(r_1, \dots, r_s) = 1 - G(1 - r_1, \dots, 1 - r_s)$$

for every $(r_1, \dots, r_s) \in [0, 1]^s$ and for every $s \in \mathbb{N}$.

All these properties are interesting when aggregating the profile of preferences into a joint one: symmetry implies that all the elements in the profile are given the same

weight; idempotency means that if all the preference degrees equal r , the final preference degree should also equal r ; monotonicity assures that if we increase all the values in the profile of preferences, the final value should also increase; and self-duality preserves the idea behind the notion of probabilistic relation in Definition 2.7, since for a self-dual aggregation function G , $G(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}^t) + G(\mathbb{Q}_{\mathcal{Y},\mathcal{X}}) = 1$. If in addition G is symmetric, we obtain that $G(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) + G(\mathbb{Q}_{\mathcal{Y},\mathcal{X}}) = 1$.

This last property means that, when G is a self-dual and symmetric aggregation function, Equation (4.10) is equivalent to $G(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) \geq G(\mathbb{Q}_{\mathcal{Y},\mathcal{X}})$.

The relations \succeq_{SP_i} , for $i = 1, \dots, 6$, can all be expressed by means of an aggregation function, as we summarise in the following proposition. Its proof is immediate and therefore omitted.

Proposition 4.72 *Let $\mathcal{X} = \{X_1, \dots, X_n\}$, $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ be two finite sets of random variables taking values on an ordered space Ω' . Then for any $i = 1, \dots, 6$ $\mathcal{X} \succeq_{\text{SP}_i} \mathcal{Y}$ if and only if it is G_i -statistically preferred to \mathcal{Y} , where the aggregation functions G_i are given by:*

$$\begin{aligned} G_1(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) &:= \min_{i,j} Q(X_i, Y_j). \\ G_2(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) &:= \max_{i=1,\dots,n} \min_{j=1,\dots,m} Q(X_i, Y_j). \\ G_3(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) &:= \min_{j=1,\dots,m} \max_{i=1,\dots,n} Q(X_i, Y_j). \\ G_4(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) &:= \max_{i,j} Q(X_i, Y_j). \\ G_5(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) &:= \max_{j=1,\dots,m} \min_{i=1,\dots,n} Q(X_i, Y_j). \\ G_6(\mathbb{Q}_{\mathcal{X},\mathcal{Y}}) &:= \min_{i=1,\dots,n} \max_{j=1,\dots,m} Q(X_i, Y_j). \end{aligned}$$

It is not difficult to see that all the aggregation functions G_i above are monotonic and comply with the boundary conditions $G_i(0, \dots, 0) = 0$ and $G_i(1, \dots, 1) = 1$. On the other hand, only G_1 and G_4 are symmetric, and none of them is self-dual.

We can also use these aggregation functions to deduce the relationships between the different conditions established in Proposition 4.3 in the case of statistical preference: it suffices to take into account that $G_1 \leq G_2 \leq G_3 \leq G_4$ and $G_1 \leq G_5 \leq G_6 \leq G_4$.

Remark 4.73 *Proposition 4.72 helps to verify each of the conditions \succeq_{SP_i} , $i = 1, \dots, 6$ by looking at the profile of preferences $\mathbb{Q}_{\mathcal{X},\mathcal{Y}}$ given by Equation (4.9):*

- $\mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y}$ if and only if all elements in the matrix are greater than or equal to $\frac{1}{2}$.

- $\mathcal{X} \succeq_{\text{SP}_2} \mathcal{Y}$ if and only if there is a row whose elements are all greater than or equal to $\frac{1}{2}$.
- $\mathcal{X} \succeq_{\text{SP}_3} \mathcal{Y}$ if and only if in each column there is at least one element greater than or equal to $\frac{1}{2}$.
- $\mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y}$ if and only if there is an element greater than or equal to $\frac{1}{2}$.
- $\mathcal{X} \succeq_{\text{SP}_5} \mathcal{Y}$ if and only if there is a column whose elements are all greater than or equal to $\frac{1}{2}$.
- $\mathcal{X} \succeq_{\text{SP}_6} \mathcal{Y}$ if and only if in each row there is at least one element greater than or equal to $\frac{1}{2}$.

See the comments after Proposition 4.3 for a related idea. \blacklozenge

The above remarks suggest that other preference relationships may be defined by means of other aggregation functions G , and this would allow us to take all the elements of the profile of preferences into account, instead of focusing on the best or worst scenarios only. Next, we explore briefly one of these possibilities: the arithmetic mean G_{mean} , given by

$$\begin{aligned} G_{\text{mean}} : \cup_{s \in \mathbb{N}} [0, 1]^s &\rightarrow [0, 1] \\ (r_1, \dots, r_s) &\mapsto \frac{r_1 + \dots + r_s}{s}. \end{aligned}$$

This is a symmetric, monotone, idempotent and self-dual aggregation function. For clarity, when \mathcal{X} is G_{mean} -statistically preferred to \mathcal{Y} we shall denote it $\mathcal{X} \succeq_{\text{SP}_{\text{mean}}} \mathcal{Y}$. The connection between $\succeq_{\text{SP}_{\text{mean}}}$ and \succeq_{SP_i} , $i = 1, \dots, 6$ is a consequence of the following result:

Proposition 4.74 *Given two finite sets of random variables \mathcal{X} and \mathcal{Y} , $\mathcal{X} = \{X_1, \dots, X_n\}$ and $\mathcal{Y} = \{Y_1, \dots, Y_m\}$, and a monotone and idempotent aggregation function G ,*

$$\mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y} \Rightarrow \mathcal{X} \succeq_{\text{SP}_G} \mathcal{Y} \Rightarrow \mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y}.$$

Proof: On the one hand, assume that $\mathcal{X} \succeq_{\text{SP}_1} \mathcal{Y}$. Then, $Q(X, Y) \geq \frac{1}{2}$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. Since G is monotone and idempotent, $G(Q_{\mathcal{X}, \mathcal{Y}}) \geq G(\frac{1}{2}, \dots, \frac{1}{2}) = \frac{1}{2}$, and consequently $\mathcal{X} \succeq_{\text{SP}_G} \mathcal{Y}$.

On the other hand, assume ex-absurdo that $G(Q_{\mathcal{X}, \mathcal{Y}}) \geq \frac{1}{2}$ and that $\mathcal{X} \not\succeq_{\text{SP}_4} \mathcal{Y}$, so that $Q(X, Y) < \frac{1}{2}$ for every $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. Then $G(Q_{\mathcal{X}, \mathcal{Y}}) \leq \max_{i,j} Q(X_i, Y_j) < \frac{1}{2}$, a contradiction. Hence, $\mathcal{X} \succeq_{\text{SP}_4} \mathcal{Y}$. \blacksquare

In particular, we see that $\succeq_{\text{SP}_{\text{mean}}}$ is an intermediate notion between \succeq_{SP_1} and \succeq_{SP_4} . To see that it is not related to \succeq_{SP_i} for $i = 2, 3, 5, 6$, consider the following example:

Example 4.75 Consider $\Omega = \{\omega_1, \omega_2\}$ ($P(\{\omega_i\}) = 1/2$), and the sets of random variables $\mathcal{X} = \{X_1, X_2, X_3\}$ and $\mathcal{Y} = \{Y\}$ defined by:

	ω_1	ω_2
X_1	0	2
X_2	0	0
X_3	2	2
Y	1	1

Then,

$$Q_{\mathcal{X}, \mathcal{Y}} := \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix} \text{ and } Q_{\mathcal{Y}, \mathcal{X}} := \begin{pmatrix} \frac{1}{2} & 1 & 0 \end{pmatrix}$$

whence Remark 4.73 implies that $\mathcal{X} \succ_{\text{SP}_i} \mathcal{Y}$, for $i = 2, 3$, and $\mathcal{Y} \succ_{\text{SP}_i} \mathcal{X}$, for $i = 5, 6$. On the other hand,

$$\frac{Q(X_1, Y) + Q(X_2, Y) + Q(X_3, Y)}{3} = \frac{1}{2},$$

and consequently $\mathcal{X} \equiv_{\text{SP}_{\text{mean}}} \mathcal{Y}$. Hence, $\mathcal{X} \succeq_{\text{SP}_{\text{mean}}} \mathcal{Y} \not\Rightarrow \mathcal{X} \succeq_{\text{SP}_i} \mathcal{Y}$ for $i = 5, 6$, and $\mathcal{Y} \succeq_{\text{SP}_{\text{mean}}} \mathcal{X} \not\Rightarrow \mathcal{Y} \succeq_{\text{SP}_i} \mathcal{X}$ for $i = 2, 3$. By comparing $\mathcal{Z}_1 = \{X_2, Y\}$ and $\mathcal{Z}_2 = \{X_3, Y\}$ with \mathcal{X} , we can see that: $\mathcal{Z}_1 \equiv_{\text{SP}_{5,6}} \mathcal{X} \succ_{\text{SP}_{\text{mean}}} \mathcal{Z}_1$ and $\mathcal{Z}_2 \equiv_{\text{SP}_{2,3}} \mathcal{X} \prec_{\text{SP}_{\text{mean}}} \mathcal{Z}_2$. Then, there are not general relationships between $\succeq_{\text{SP}_{\text{mean}}}$ and \succeq_{SP_i} for $i = 2, 3, 5, 6$. \blacklozenge

4.2 Modelling imprecision in decision making problems

In this section, we shall show how the above results can be applied in two different scenarios where imprecision enters a decision problem: the case where we have imprecise information about the utilities of the different alternatives, and that where we have imprecise beliefs about the states of nature.

4.2.1 Imprecision on the utilities

Let us start with the first case. Consider a decision problem where we must choose between two alternatives X and Y whose respective utilities depend on the values ω of the states of nature. Assume that we have precise information about the probabilities of these states of nature, so that X and Y can be seen as random variables defined on a probability space (Ω, \mathcal{A}, P) . If we have imprecise knowledge about the utilities $X(\omega)$ associated with the different states of nature, one possible model would be to associate to any $\omega \in \Omega$ a set $\Gamma(\omega)$ that is sure to include the ‘true’ utility $X(\omega)$. By doing this, we obtain a multi-valued mapping $\Gamma : \Omega \rightarrow \mathcal{P}(\Omega')$, and all we know about X is that it is one of the measurable selections of Ω , that were defined in Equation (2.21) by:

$$S(\Gamma) = \{U : \Omega \rightarrow \Omega' \text{ r.v.} : U(\omega) \in \Gamma(\omega) \text{ for every } \omega \in \Omega\}. \quad (4.11)$$

In this paper, we shall consider only multi-valued mappings satisfying the measurability condition:

$$\Gamma^*(A) := \{\omega \in \Omega : \Gamma(\omega) \cap A \neq \emptyset\} \in \mathcal{A} \text{ for any } A \in \mathcal{A}'.$$

As we saw in Definition 2.42, these multi-valued mappings are called random sets.

Our comparison of two alternatives with imprecise utilities results thus in the comparison of two random sets Γ_1, Γ_2 , that we shall make by means of their respective sets of measurable selections $S(\Gamma_1), S(\Gamma_2)$ determined by Equation (2.21). For simplicity, we shall use the notation $\Gamma_1 \succeq \Gamma_2$ instead of $S(\Gamma_1) \succeq S(\Gamma_2)$ when no confusion is possible.

Let us begin by studying the comparison of random sets by means of stochastic dominance.

Proposition 4.76 *Let (Ω, \mathcal{A}, P) be a probability space, $(\Omega', \mathcal{P}(\Omega'))$ a measurable space, with Ω' a finite subset of \mathbb{R} , and Γ_X, Γ_Y be two random sets. The following equivalences hold:*

- (a) $\Gamma_X \succeq_{\text{FSD}_1} \Gamma_Y \Leftrightarrow \min \Gamma_X \succeq_{\text{FSD}} \max \Gamma_Y.$
- (b) $\Gamma_X \succeq_{\text{FSD}_2} \Gamma_Y \Leftrightarrow \Gamma_X \succeq_{\text{FSD}_3} \Gamma_Y \Leftrightarrow \max \Gamma_X \succeq_{\text{FSD}} \max \Gamma_Y.$
- (c) $\Gamma_X \succeq_{\text{FSD}_4} \Gamma_Y \Leftrightarrow \max \Gamma_X \succeq_{\text{FSD}} \min \Gamma_Y.$
- (d) $\Gamma_X \succeq_{\text{FSD}_5} \Gamma_Y \Leftrightarrow \Gamma_X \succeq_{\text{FSD}_6} \Gamma_Y \Leftrightarrow \min \Gamma_X \succeq_{\text{FSD}} \min \Gamma_Y.$

Proof: The result follows from Proposition 4.19, taking into account that given a random set Γ taking values on a finite space, the lower distribution function associated with its set $S(\Gamma)$ of measurable selections is induced by $\max \Gamma$ and its upper distribution function is induced by $\min \Gamma$. ■

Moreover, we can characterise the conditions $\succeq_{\text{FSD}_i}, i = 1, \dots, 6$ even for random sets that take values on infinite spaces. To see how this comes out, we shall consider the upper and lower probabilities induced by the random set. Recall that, from Equation (2.22), they are defined by:

$$\begin{aligned} P^*(A) &= P(\{\omega : \Gamma(\omega) \cap A \neq \emptyset\}) \text{ and} \\ P_*(A) &= P(\{\omega : \emptyset \neq \Gamma(\omega) \subseteq A\}) \end{aligned}$$

for any $A \in \mathcal{A}'$. As we have already seen in Equation (2.24), the upper and lower probabilities of a random set constitute upper and lower bounds of the probabilities induced by the measurable selections:

$$P_*(A) \leq P_U(A) \leq P^*(A) \quad \forall U \in S(\Gamma),$$

and in particular their associated cumulative distributions provide lower and upper bounds of the lower and upper distribution functions associated with $S(\Gamma)$.

We have seen in Theorem 2.46 that when $P^*(A)$ is attained by the probabilities induced by the measurable selections for any element $A \in \mathcal{A}$, the supremum and infimum of the integrals of a gamble with respect to the measurable selections can be expressed by means of the Choquet integral of the gamble with respect to P^* and P_* . This result allows to characterise the imprecise stochastic dominance between random sets by means of the comparison of Choquet or Aumann integrals. Recall that we have denoted by \mathcal{U}^* the set of increasing and bounded functions $u : [0, 1] \rightarrow \mathbb{R}$.

Proposition 4.77 *Let (Ω, \mathcal{A}, P) be a probability space. Consider the measurable space $([0, 1], \beta_{[0,1]})$ and let $\Gamma_X, \Gamma_Y : \Omega \rightarrow \mathcal{P}([0, 1])$ be two random sets. If for all $A \in \beta_{[0,1]}$ it holds that $P_X^*(A) = \max \mathcal{P}(\Gamma_X)(A)$ and $P_Y^*(A) = \max \mathcal{P}(\Gamma_Y)(A)$, the following equivalences hold:*

1. $\Gamma_X \succeq_{\text{FSD}_1} \Gamma_Y \Leftrightarrow (C) \int udP_{X^*} \geq (C) \int udP_Y^*$ for every $u \in \mathcal{U}^*$.
2. $\Gamma_X \succeq_{\text{FSD}_2} \Gamma_Y \Rightarrow (C) \int udP_X^* \geq (C) \int udP_Y^*$ for every $u \in \mathcal{U}^*$.
3. $\Gamma_X \succeq_{\text{FSD}_3} \Gamma_Y \Rightarrow (C) \int udP_X^* \geq (C) \int udP_X^*$ for every $u \in \mathcal{U}^*$.
4. $\Gamma_X \succeq_{\text{FSD}_4} \Gamma_Y \Rightarrow (C) \int udP_X^* \geq (C) \int udP_{X^*}$ for every $u \in \mathcal{U}^*$.
5. $\Gamma_X \succeq_{\text{FSD}_5} \Gamma_Y \Rightarrow (C) \int udP_{X^*} \geq (C) \int udP_{X^*}$ for every $u \in \mathcal{U}^*$.
6. $\Gamma_X \succeq_{\text{FSD}_6} \Gamma_Y \Rightarrow (C) \int udP_{X^*} \geq (C) \int udP_{Y^*}$ for every $u \in \mathcal{U}^*$.

Proof: Consider $u \in \mathcal{U}^*$. We deduce from Theorem 2.46 that, under the hypotheses of the proposition,

$$(C) \int udP_X^* = \sup_{U \in S(\Gamma_X)} \int udP_U = \overline{E}_{S(\Gamma_X)}(u) \text{ and}$$

$$(C) \int udP_{X^*} = \inf_{U \in S(\Gamma_X)} \int udP_U = \underline{E}_{S(\Gamma_X)}(u)$$

and similarly:

$$(C) \int udP_Y^* = \sup_{U \in S(\Gamma_Y)} \int udP_U = \overline{E}_{S(\Gamma_Y)}(u) \text{ and}$$

$$(C) \int udP_{Y^*} = \inf_{U \in S(\Gamma_Y)} \int udP_U = \underline{E}_{S(\Gamma_Y)}(u)$$

The result follows then applying Theorem 4.23. ■

Let us discuss next the comparison of random sets by means of statistical preference. When the utility space Ω' is finite, we obtain a result related to Proposition 4.76:

Proposition 4.78 *Let (Ω, \mathcal{A}, P) be a probability space, $(\Omega', \mathcal{P}(\Omega'))$ a measurable space, with Ω' finite, and Γ_X, Γ_Y be two random sets. The following equivalences hold:*

- (a) $\Gamma_X \succeq_{SP_1} \Gamma_Y \Leftrightarrow \min \Gamma_X \succeq_{SP} \max \Gamma_Y$.
- (b) $\Gamma_X \succeq_{SP_2} \Gamma_Y \Leftrightarrow \Gamma_X \succeq_{SP_3} \Gamma_Y \Leftrightarrow \max \Gamma_X \succeq_{SP} \max \Gamma_Y$.
- (c) $\Gamma_X \succeq_{SP_4} \Gamma_Y \Leftrightarrow \max \Gamma_X \succeq_{SP} \min \Gamma_Y$.
- (d) $\Gamma_X \succeq_{SP_5} \Gamma_Y \Leftrightarrow \Gamma_X \succeq_{SP_6} \Gamma_Y \Leftrightarrow \min \Gamma_X \succeq_{SP} \min \Gamma_Y$.

Proof: The result follows from Proposition 4.7, taking into account that statistical preference satisfies the monotonicity condition of Equation (4.2) and that if Γ is a random set taking values on a finite space, then the mappings $\min \Gamma, \max \Gamma$ belong to $S(\Gamma)$. ■

In particular, we deduce that we can focus on the minimum and maximum measurable selections in order to characterise these extensions of statistical preference.

Corollary 4.79 *Let (Ω, \mathcal{A}, P) be a probability space, Ω' a finite space and consider two random sets $\Gamma_X, \Gamma_Y : \Omega \rightarrow \mathcal{P}(\Omega')$. Then for every $i = 1, \dots, 6$:*

$$\Gamma_X \succeq_{SP_i} \Gamma_Y \Leftrightarrow \{\min \Gamma_X, \max \Gamma_X\} \succeq_{SP_i} \{\min \Gamma_Y, \max \Gamma_Y\}. \quad (4.12)$$

These two results are interesting because random sets taking values on finite spaces are quite common in practice; they have been studied in detail in [59, 127], and one of their most interesting properties is that they constitute equivalent models to belief and plausibility functions [170].

Note that the equivalence in Equation (4.12) does not hold for the relation $\succeq_{SP_{\text{mean}}}$ defined in Section 4.1.2.

Example 4.80 *Consider the probability space (Ω, \mathcal{A}, P) where $\Omega = \{\omega_1, \omega_2\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ and P is a probability uniformly distributed on Ω , and let Γ_X be the random set given by $\Gamma_X(\omega_1) = \{0, 1\}, \Gamma_X(\omega_2) = \{0, 2, 3, 4\}$, and let Γ_Y be single-valued random set given by $\Gamma_Y(\omega_1) = \{1\} = \Gamma_Y(\omega_2)$. Then $\min \Gamma_X$ is the constant random variable on 0, while $\max \Gamma_X$ is given by $\max \Gamma_X(\omega_1) = 1, \max \Gamma_X(\omega_2) = 4$. Hence, if we compare the set $\{\min \Gamma_X, \max \Gamma_X\}$ with Γ_Y by means of $\succeq_{SP_{\text{mean}}}$ we obtain*

$$\frac{Q(\min \Gamma_X, \Gamma_Y) + Q(\max \Gamma_X, \Gamma_Y)}{2} = \frac{0 + 0.75}{2} = 0.375$$

and thus $\Gamma_Y \succ_{\text{SP}_{\text{mean}}} \{\min \Gamma_X, \max \Gamma_X\}$. On the other hand, the set of selections of Γ_X is given by (where a selection X is identified with the vector $(U(\omega_1), U(\omega_2))$):

$$S(\Gamma_X) = \{(0, 0), (0, 2), (0, 3), (0, 4), (1, 0), (1, 2), (1, 3), (1, 4)\},$$

from which we deduce that $\Gamma_X \succ_{\text{SP}_{\text{mean}}} \Gamma_Y$. \blacklozenge

4.2.2 Imprecision on the beliefs

We next consider the case where we want to choose between two random variables X and Y defined from Ω to Ω' , and there is some uncertainty about the probability distribution P of the different states of nature $\omega \in \Omega$, that we model by means of a set \mathcal{P} of probability distributions on Ω . Then we may associate with X a set \mathcal{X} of random variables, that correspond to the transformations of X under any of the probability distributions in \mathcal{P} ; and similarly for Y . We end up thus with two sets \mathcal{X}, \mathcal{Y} of random variables, and we should establish methods to determine which of these two sets is preferable.

One particular case where this situation may arise is in the context of missing data [218]. We may divide the variables determining the states of nature in two groups: one for which we have precise information, that we model by means of a probability measure P over the different states, and another about which are completely ignorant, knowing only which are the different states, but nothing more. Then we may get to the classical scenario by fixing the value of the variables in this second group: for each of these values the alternatives may be seen as random variables, using the probability measure P to determine the probabilities of the different rewards. Hence, by doing this we would transform the two alternatives X and Y into two sets of alternatives \mathcal{X}, \mathcal{Y} , considering all the possible values of the variables in the second group.

In this situation, we may compare the sets \mathcal{X}, \mathcal{Y} by means of the generalisations of statistical preference or stochastic dominance we have discussed in Section 4.1; however, we argue that other notions may make more sense in this context. This is because conditions $\succeq_1, \dots, \succeq_6$ are based on considering a particular pair (X_1, Y_1) in $\mathcal{X} \times \mathcal{Y}$ and on comparing X_1 with Y_1 by means of the binary relation \succeq . However, any X_1 in \mathcal{X} corresponds to a particular choice of a probability measure $P \in \mathcal{P}$, and similarly for any $Y_1 \in \mathcal{Y}$; and if we use an epistemic interpretation of our uncertainty under which only one $P \in \mathcal{P}$ is the ‘true’ model, it makes no sense to compare X_1 and Y_1 based on a different distribution. This is particularly clear in case we want to apply statistical preference, which is based on comparing $P(X > Y)$ with $P(Y > X)$, where P is the initial probability measure.

To make this explicit, in this section we may denote our sets of alternatives by $\mathcal{X} := \{(X, P) : P \in \mathcal{P}\}$ and $\mathcal{Y} := \{(Y, P) : P \in \mathcal{P}\}$, meaning that our utilities are precise (and are determined by the variables X and Y , respectively), while our beliefs are imprecise and are modelled by the set \mathcal{P} . To avoid confusions, we will now write

$X \succeq^P Y$ to express that X is preferred to Y when we consider the probability measure P in the initial probability space. Then we can establish the following definitions:

Definition 4.81 *Let \succeq be a binary relation on random variables. We say that:*

- \mathcal{X} is strongly \mathcal{P} preferred to \mathcal{Y} , and denote it $\mathcal{X} \succeq_s^{\mathcal{P}} \mathcal{Y}$, when $X \succeq^P Y$ for every $P \in \mathcal{P}$;
- \mathcal{X} is weakly \mathcal{P} preferred, and denote it $\mathcal{X} \succeq_w^{\mathcal{P}} \mathcal{Y}$, to \mathcal{Y} when $X \succeq^P Y$ for some $P \in \mathcal{P}$.

Obviously, the strong preference implies the weak one. To see that they are not equivalent, consider the following simple example:

Example 4.82 *Let \succeq be the binary relation associated with statistical preference and consider the variables X, Y that represent the results of the dices A and B , respectively, in Example 3.83. If we consider the uniform distribution P_1 in all the die outcomes, we obtain $Q(X, Y) = \frac{5}{9}$, so that $X \succeq_{\text{SP}}^{P_1} Y$; if we take instead the uniform distribution P_2 on $\{1, 2, 3\}$, then $Q(X, Y) = \frac{1}{9}$, and as a consequence $Y \succeq_{\text{SP}}^{P_2} X$. Hence, X is weakly $\{P_1, P_2\}$ statistically preferred to Y , but not strongly so. \blacklozenge*

With respect to the notions established in Section 4.1, it is not difficult to establish the following result. Its proof is immediate, and therefore omitted.

Proposition 4.83 *Let \mathcal{X}, \mathcal{Y} be the sets of alternatives considered above, and let \succeq be a binary relation. Then*

$$\mathcal{X} \succeq_1 \mathcal{Y} \Rightarrow \mathcal{X} \succeq_s^{\mathcal{P}} \mathcal{Y} \Rightarrow \mathcal{X} \succeq_w^{\mathcal{P}} \mathcal{Y} \Rightarrow \mathcal{X} \succeq_4 \mathcal{Y}.$$

To see that the converse implications do not hold, consider the following example:

Example 4.84 *Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$, the set of probabilities*

$$\mathcal{P} := \{P : P(\omega_1) > P(\omega_2), P(\omega_2) \in [0, 0.2]\}$$

and the alternatives X, Y given by

	ω_1	ω_2	ω_3
X	1	0	1
Y	0	1	1

If we consider the sets $\mathcal{X} = \{(X, P) : P \in \mathcal{P}\}$ and $\mathcal{Y} = \{(Y, P) : P \in \mathcal{P}\}$ and we compare them by means of stochastic dominance, it is clear that $\mathcal{X} \succ_s^{\mathcal{P}} \mathcal{Y}$; however, it does not

hold that $\mathcal{X} \succeq_{\text{FSD}_1} \mathcal{Y}$: if we consider $P_1 := (0.3, 0.2, 0.5)$ and $P_2 := (0.1, 0, 0.9)$, it holds that $(Y, P_2) \succ_{\text{FSD}} (X, P_1)$.

Moreover, in this example we also have that \mathcal{X} is strictly weakly \mathcal{P} -preferred to \mathcal{Y} while $\mathcal{X} \equiv_{\text{FSD}_4} \mathcal{Y}$. \blacklozenge

Remark 4.85 If the binary relation \succeq we start with is complete, so is the weak \mathcal{P} -preference. In that case, we obtain that $\mathcal{X} \succ_w^{\mathcal{P}} \mathcal{Y}$ implies that $\mathcal{X} \succ_s^{\mathcal{P}} \mathcal{Y}$, because if $\mathcal{X} \succ_w^{\mathcal{P}} \mathcal{Y}$ we must have that $(X, P) \succ (Y, P)$ for every $P \in \mathcal{P}$.

Moreover, when $\mathcal{X} \equiv_w^{\mathcal{P}} \mathcal{Y}$, we may have strict preference, indifference or incomparability with respect to strong \mathcal{P} -preference. \blacklozenge

In what follows, we study in some detail the notions of weak and strong preference for particular choices of the binary relation \succeq . If \succeq corresponds to expected utility, strong preference of \mathcal{X} over \mathcal{Y} means that X is preferred to Y with respect to all the probability measures P in \mathcal{P} , and then it is related to the idea of *maximality* [205]; on the other hand, weak preference means that X is preferred to Y (i.e., it is the optimal alternative) with respect to some of the elements of \mathcal{P} ; this idea is close to the criterion of E -admissibility [107]. See also Remark 4.13 and [43, Section 3.2].

When \succeq is the binary relation associated with stochastic dominance, we obtain the following.

Proposition 4.86 Consider a set \mathcal{P} of probability measures on Ω , and let X, Y be two real-valued random variables on Ω . Let us define the sets $\mathcal{F}_X := \{F_X^P : P \in \mathcal{P}\}$ and $\mathcal{F}_Y := \{F_Y^P : P \in \mathcal{P}\}$.

1. $\overline{F}_X \leq \underline{F}_Y \Rightarrow \mathcal{X}$ is strongly \mathcal{P} -preferred to \mathcal{Y} with respect to stochastic dominance.
2. \mathcal{X} is weakly \mathcal{P} -preferred to \mathcal{Y} with respect to stochastic dominance $\Rightarrow \underline{F}_X \leq \overline{F}_Y$.

Proof: Assume that $\overline{F}_X \leq \underline{F}_Y$. Then, for any $P \in \mathcal{P}$ it holds that:

$$F_X^P \leq \overline{F}_X \leq \underline{F}_Y \leq F_Y^P.$$

Then, \mathcal{X} is strongly \mathcal{P} -preferred to \mathcal{Y} with respect to first degree stochastic dominance.

Now, assume that \mathcal{X} is weakly \mathcal{P} -preferred to \mathcal{Y} with respect to first degree stochastic dominance. Then there exists $P \in \mathcal{P}$ such that $F_X^P \leq F_Y^P$. Then, in particular, $\mathcal{X} \succeq_{\text{FSD}_4} \mathcal{Y}$, and by Proposition 4.19 we deduce that $\underline{F}_X \leq \overline{F}_Y$. \blacksquare

Note that this result could also be derived from Propositions 4.19 and 4.83.

Finally, when \succeq corresponds to statistical preference, we can apply Remark 4.85, because \succeq is a complete relation. In addition, we can establish the following result:

Proposition 4.87 Consider a set \mathcal{P} of probability measures, and let \underline{P}, \bar{P} denote its lower and upper envelopes, given by Equation (2.18). Let X, Y be two real-valued random variables on Ω , and let $u = I_{(0,+\infty)} - I_{(-\infty,0)}$.

1. X is strongly \mathcal{P} statistically preferred to $Y \Leftrightarrow \underline{P}(u(X - Y)) \geq 0$.
2. X is weakly \mathcal{P} statistically preferred to $Y \Rightarrow \bar{P}(u(X - Y)) \geq 0$. The converse holds if $\mathcal{P} = \mathcal{M}(\bar{P})$.

Proof: The result follows simply by considering that if X, Y are random variables on a probability space (Ω, \mathcal{A}, P) , then, by applying Equation (3.1), $X \succeq_{\text{SP}}^P Y$ if and only if $P(u(X - Y)) \geq 0$, where we also use P to denote the expectation operator associated with the probability measure P .

To see that the converse of the second statement holds when $\mathcal{P} = \mathcal{M}(\underline{P})$, note that the upper envelope \bar{P} of \mathcal{P} is a coherent lower prevision. From [205, Section 3.3.3], given the bounded random variable $u(X - Y)$ there exists a probability P in $\mathcal{M}(\underline{P})$ such that $P(u(X - Y)) = \bar{P}(u(X - Y))$. ■

The above result can be related to the lower median, as in [46, 148]. For this, let us define the *lower median* of $X - Y$ by the credal set $\mathcal{M}(\underline{P})$ by

$$\underline{Me}(X - Y) := \inf\{Me_P(X - Y) : P \in \mathcal{M}(\underline{P})\},$$

and its *upper median* by

$$\bar{Me}(X - Y) := \sup\{Me_P(X - Y) : P \in \mathcal{M}(\underline{P})\},$$

where $Me_P(X - Y)$ denotes the median of $X - Y$ when P is the probability of the initial space.

Then, we deduce from Proposition 4.64 that

$$\underline{Me}(X - Y) > 0 \Rightarrow X \succeq_{\text{SP},s}^{\mathcal{M}(\underline{P})} Y \Rightarrow \bar{Me}(X - Y) \geq 0,$$

and that

$$\bar{Me}(X - Y) > 0 \Rightarrow X \succeq_{\text{SP},w}^{\mathcal{M}(\underline{P})} Y \Rightarrow \underline{Me}(X - Y) \geq 0.$$

A related result was established in [46, Proposition 4], by means of a slightly different definition of median. See also Proposition 4.64, and [83, 164] for approaches based on the expected utility model.

4.3 Modelling the joint distribution in an imprecise framework

Statistical preference is a stochastic order that depends on the joint distribution of the random variables. This joint distribution function can be determined, according

to Sklar's Theorem (Theorem 2.27), from the marginals by means of a copula. In the imprecise context we are dealing with in this chapter, there may be imprecision either in the marginal distribution functions or in the copula that links the marginals. In the former case, we can model the lack of information by means of p-boxes, and in the second one the should consider a set of possible copulas. In both situations we shall obtain a set of bivariate distribution functions.

In order to determine the mathematical model for this situation, we shall consider two steps: on the one hand, we shall study how to model sets of bivariate distribution functions, since the lower and upper bounds are not, in general, distribution functions. To deal with this problem, we shall extend the notion of p-box when considering bivariate distribution functions, and we will investigate under which conditions such bivariate p-box can define a coherent lower probability. Then, we shall consider two marginal imprecise distribution functions and we will try to build from them a joint distribution. In this context, the main result is to extend Sklar's Theorem to an imprecise framework; we shall also study the application of these results can be applied into bivariate stochastic orders.

4.3.1 Bivariate distribution with imprecision

Bivariate p-boxes

Let Ω_1, Ω_2 be two totally ordered spaces. As in [198], we assume without loss of generality that both have a maximum element, that we denote respectively by x^*, y^* . Note that this is trivial in the case of finite spaces.

We start by introducing standardized functions and bivariate distribution functions.

Definition 4.88 *Consider two ordered spaces Ω_1, Ω_2 . A map $F : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ is called standardized when it is component-wise increasing and $F(x^*, y^*) = 1$. It is called a distribution function when moreover it satisfies the rectangle inequality:*

$$(RI) : F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \geq 0$$

for every $x_1, x_2 \in \Omega_1$ and $y_1, y_2 \in \Omega_2$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$.

Here, and in what follows, we shall make an assumption of *logical independence*, meaning that we consider all values in the product space $\Omega_1 \times \Omega_2$ to be possible.

The rectangle inequality is equivalent to monotonicity in the univariate case, so in that case a distribution function is simply an increasing and normalized function $F : \mathcal{X} \rightarrow [0, 1]$. Moreover, a lower envelope of univariate distribution functions is again a distribution function, by Proposition 2.34. Unfortunately, the situation is not as clear

cut in the bivariate case: the envelopes of a set of distribution functions are standardized maps, but not necessarily distribution functions.

Proposition 4.89 *Let Ω_1 and Ω_2 be two ordered spaces and \mathcal{F} be a family of distribution functions $F : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$. Their lower and upper envelopes $\underline{F}, \overline{F} : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$, given by*

$$\underline{F}(x, y) = \inf_{F \in \mathcal{F}} F(x, y) \text{ and } \overline{F}(x, y) = \sup_{F \in \mathcal{F}} F(x, y)$$

for every $x \in \Omega_1, y \in \Omega_2$, are standardized maps.

Proof: It suffices to take into account that the monotonicity and normalization properties are preserved by lower and upper envelopes. ■

To see that these envelopes are not necessarily distribution functions, consider the following example:

Example 4.90 *Take $\Omega_1 = \Omega_2 = \{a, b, c\}$, with $a < b < c$ and let F_1, F_2 be the distribution functions determined by the following joint probability measures:*

X_1, Y_1	a	b	c	X_2, Y_2	a	b	c
a	0.1	0.1	0	a	0.4	0	0.2
b	0.4	0.1	0	b	0.1	0	0
c	0	0	0.3	c	0.1	0	0.2

Then F_1 and F_2 are given by:

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
F_1	0.1	0.2	0.2	0.5	0.7	0.7	0.5	0.7	1
F_2	0.4	0.4	0.6	0.5	0.5	0.7	0.6	0.6	1

and their lower and upper envelopes are given by:

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
\underline{F}	0.1	0.2	0.2	0.5	0.5	0.7	0.5	0.6	1
\overline{F}	0.4	0.4	0.6	0.5	0.7	0.7	0.6	0.7	1

Then

$$\underline{F}(b, b) + \underline{F}(a, a) - \underline{F}(a, b) - \underline{F}(b, a) = 0.5 + 0.1 - 0.2 - 0.5 = -0.1 < 0$$

and

$$\overline{F}(b, c) + \overline{F}(a, b) - \overline{F}(a, c) - \overline{F}(b, b) = 0.7 + 0.4 - 0.6 - 0.7 = -0.2 < 0.$$

As a consequence, neither \underline{F} nor \overline{F} are distribution functions. ♦

Taking this result into account, we give the following definition:

Definition 4.91 Consider two ordered spaces Ω_1, Ω_2 , and let $\underline{F}, \overline{F} : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ be two standardized functions satisfying $\underline{F}(x, y) \leq \overline{F}(x, y)$ for every $x \in \Omega_1, y \in \Omega_2$. Then the pair $(\underline{F}, \overline{F})$ is called a bivariate p-box.

Proposition 4.89 shows that bivariate p-boxes can be obtained in particular by means of a set of distribution functions, taking lower and upper envelopes. However, not all bivariate p-boxes are of this type: if we consider for instance a map $\underline{F} = \overline{F}$ that is standardized but not a distribution function, then there is no bivariate distribution function between \underline{F} and \overline{F} , and as a consequence these cannot be obtained as envelopes of a set of distribution functions. Our next paragraph will deepen into this matter, by means of the notion of coherence of lower probabilities. In particular, we shall investigate how Theorem 2.35 could be extended to bivariate p-boxes.

Lower probabilities and p-boxes

In order to define a lower probability from a bivariate p-box, let us now introduce a notation similar to the one of Section 2.2.1.

Consider two ordered spaces Ω_1, Ω_2 , and let $(\underline{F}, \overline{F})$ be a bivariate p-box on $\Omega_1 \times \Omega_2$. Denote

$$A_{(x,y)} := \{(x', y') \in \Omega_1 \times \Omega_2 : x' \leq x, y' \leq y\},$$

and let us define

$$\mathcal{K}_1 := \{A_{(x,y)} : x \in \Omega_1, y \in \Omega_2\} \text{ and } \mathcal{K}_2 := \{A_{(x,y)}^c : x \in \Omega_1, y \in \Omega_2\}.$$

The maps \underline{F} and \overline{F} can be used to define the lower probabilities $\underline{P}_{\underline{F}} : \mathcal{K}_1 \rightarrow \mathbb{R}$ and $\underline{P}_{\overline{F}} : \mathcal{K}_2 \rightarrow \mathbb{R}$ by:

$$\underline{P}_{\underline{F}}(A_{(x,y)}) = \underline{F}(x, y) \quad \text{and} \quad \underline{P}_{\overline{F}}(A_{(x,y)}^c) = 1 - \overline{F}(x, y). \quad (4.13)$$

Define now $\mathcal{K} := \mathcal{K}_1 \cup \mathcal{K}_2$; note that $A_{(x^*, y^*)} = \Omega_1 \times \Omega_2$, where x^*, y^* are the maximum of Ω_1 and Ω_2 , respectively. Thus, both $\Omega_1 \times \Omega_2$ and \emptyset belong to \mathcal{K} .

Definition 4.92 The lower probability induced by $(\underline{F}, \overline{F})$ is the map $\underline{P}_{(\underline{F}, \overline{F})} : \mathcal{K} \rightarrow [0, 1]$ given by:

$$\underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}) = \underline{F}(x, y), \quad \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}^c) = 1 - \overline{F}(x, y) \quad (4.14)$$

for every $x \in \Omega_1, y \in \Omega_2$.

Note that $\underline{P}_{(\underline{F}, \overline{F})}(\Omega_1 \times \Omega_2) = 1$ and $\underline{P}_{(\underline{F}, \overline{F})}(\emptyset) = 0$ because \underline{F} and \overline{F} are standardized.

In this section, we are going to study which properties of the lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ can be characterised in terms of the lower and upper distribution functions \underline{F} and \overline{F} .

Avoiding sure loss We begin with the property of avoiding sure loss. Recall that, as we saw in Definition 2.29, a lower probability \underline{P} with domain $\mathcal{K} \subseteq \mathcal{P}(\Omega_1 \times \Omega_2)$ avoids sure loss if and only if there is a finitely additive probability $P : \mathcal{P}(\Omega_1 \times \Omega_2) \rightarrow [0, 1]$ that dominates \underline{P} on its domain. This is a consequence of [205, Corollary 3.2.3 and Theorem 3.3.3].

Proposition 4.93 *The lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ induced by the bivariate p -box $(\underline{F}, \overline{F})$ by means of Equation (4.14) avoids sure loss if and only if there is a distribution function $F : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ satisfying $\underline{F} \leq F \leq \overline{F}$.*

Proof: We begin with the direct implication. Assume that $\underline{P}_{(\underline{F}, \overline{F})}$ avoids sure loss. Then, there exists a finitely additive probability $P : \mathcal{P}(\Omega_1 \times \Omega_2) \rightarrow [0, 1]$ such that $P(A) \geq \underline{P}_{(\underline{F}, \overline{F})}(A)$ for every $A \in \mathcal{K}$. Let us define the map $F_P : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ by $F_P(x, y) = P(A_{(x, y)})$. Then F_P is a distribution function that is bounded between \underline{F} and \overline{F} :

- Consider $x_1, x_2 \in \Omega_1$ and $y_1, y_2 \in \Omega_2$ such that $x_1 \leq x_2, y_1 \leq y_2$. Then:

$$F_P(x_1, y_1) = P(A_{(x_1, y_1)}) \leq P(A_{(x_2, y_2)}) = F_P(x_2, y_2)$$

because P is monotone.

- $F_P(x^*, y^*) = P(A_{(x^*, y^*)}) = P(\Omega_1 \times \Omega_2) = 1$.
- Consider $x_1, x_2 \in \Omega_1$ and $y_1, y_2 \in \Omega_2$ such that $x_1 \leq x_2, y_1 \leq y_2$. Then it holds that

$$\begin{aligned} & F_P(x_1, y_1) + F_P(x_2, y_2) - F_P(x_1, y_2) - F_P(x_2, y_1) \\ &= P(A_{(x_1, y_1)}) + P(A_{(x_2, y_2)}) - P(A_{(x_1, y_2)}) - P(A_{(x_2, y_1)}) \\ &= P(\{(x, y) \in \Omega_1 \times \Omega_2 : x_1 < x \leq x_2, y_1 < y \leq y_2\}) \geq 0. \end{aligned}$$

- For every $x \in \Omega_1, y \in \Omega_2$,

$$F_P(x, y) = P(A_{(x, y)}) \geq \underline{P}_{(\underline{F}, \overline{F})}(A_{(x, y)}) = \underline{F}(x, y),$$

and on the other hand,

$$\begin{aligned} F_P(x, y) &= P(A_{(x, y)}) = 1 - P(A_{(x, y)}^c) \\ &\leq 1 - \underline{P}_{(\underline{F}, \overline{F})}(A_{(x, y)}^c) = 1 - (1 - \overline{F}(x, y)) = \overline{F}(x, y). \end{aligned}$$

Conversely, assume that $F : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ is a distribution function that lies between \underline{F} and \overline{F} , and let us define the finitely additive probability P_F on the field generated by \mathcal{K} by means of

$$\begin{aligned} & P_F(\{(x, y) \in \Omega_1 \times \Omega_2 : x_1 < x \leq x_2, y_1 < y \leq y_2\}) \\ &= F_P(x_1, y_1) + F_P(x_2, y_2) - F_P(x_1, y_2) - F_P(x_2, y_1) \geq 0. \end{aligned} \quad (4.15)$$

Then it follows that $P_F(A_{(x,y)}) = F(x, y) \geq \underline{F}(x, y) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)})$ and moreover $P_F(A_{(x,y)}^c) = 1 - F(x, y) \geq 1 - \overline{F}(x, y) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}^c)$.

Since any finitely additive probability on a field of events has a finitely additive extension to $\mathcal{P}(\Omega_1 \times \Omega_2)$, we deduce that there is a finitely additive probability that dominates $\underline{P}_{(\underline{F}, \overline{F})}$, and as a consequence this lower probability avoids sure loss. ■

This result allows us to focus on the lower and upper distributions of the p-box, that shall simplify search for necessary and sufficient conditions. We shall say that $(\underline{F}, \overline{F})$ avoids sure loss when the lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ it induces by means of Equation (4.14) does. Our next result gives a necessary condition:

Proposition 4.94 *If $(\underline{F}, \overline{F})$ avoids sure loss, then for every $x_1, x_2 \in \Omega_1$ and $y_1, y_2 \in \Omega_2$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$ it holds that*

$$(\mathbf{I} - \mathbf{RIO}) : \quad \overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0.$$

Proof: Assume that $(\underline{F}, \overline{F})$ avoids sure loss. By Proposition 4.93, there is a distribution function F bounded by $\underline{F}, \overline{F}$. Given $x_1, x_2 \in \Omega_1$ and $y_1, y_2 \in \Omega_2$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$, it follows from (RI) that

$$\begin{aligned} 0 &\leq F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \\ &\leq \overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1), \end{aligned}$$

where the second inequality follows from $\underline{F} \leq F \leq \overline{F}$. ■

Let us show that this necessary condition is not sufficient in general:

Example 4.95 *Consider $\Omega_1 = \Omega_2 = \{a, b, c\}$, with $a < b < c$ and let \underline{F} and \overline{F} be given by:*

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
\underline{F}	0	0.65	0.7	0.2	0.8	0.8	0.35	0.9	1
\overline{F}	0.1	0.7	0.7	0.25	0.8	0.8	0.4	0.9	1

It is immediate to check that both maps are standardized and that together they satisfy (I-RIO). However, $(\underline{F}, \overline{F})$ does not avoid sure loss: from Proposition 4.93, it suffices to show that there is no distribution function F bounded by $\underline{F}(x, y)$ and $\overline{F}(x, y)$ for every $x, y \in \{a, b, c\}$. To see that this is indeed the case, note that any distribution function $F \in (\underline{F}, \overline{F})$ should satisfy

$$F(a, c) = 0.7, \quad F(b, b) = 0.8, \quad F(b, c) = 0.8, \quad F(c, b) = 0.9 \quad \text{and} \quad F(c, c) = 1.$$

By (RI) to $(x_1, y_1) = (a, b)$ and $(x_2, y_2) = (b, c)$, we deduce that $F(a, b) = 0.7$, and then applying again the rectangle inequality we deduce that

$$F(b, b) + F(a, a) - F(a, b) - F(b, a) = 0.8 + F(a, a) - 0.7 - F(b, a) \geq 0$$

if and only if $F(a, a) + 0.1 \geq F(b, a)$, whence $F(a, a) = 0.1$ and $F(b, a) = 0.2$. If we now apply (RI) to $(x_1, y_1) = (b, a)$ and $(x_2, y_2) = (c, b)$, we deduce that

$$F(c, b) + F(b, a) - F(b, b) - F(c, a) = 0.9 + 0.2 - 0.8 - F(c, a) \geq 0$$

if and only if $F(c, a) \leq 0.3$. But on the other hand we must have $F(c, a) \geq \underline{F}(c, a) = 0.35$, a contradiction. Hence, $(\underline{F}, \overline{F})$ does not avoid sure loss. \blacklozenge

However, (I-RI0) is a necessary and sufficient condition when both Ω_1, Ω_2 are binary spaces.

Proposition 4.96 *Assume that both $\Omega_1 = \{x_1, x_2\}$ and $\Omega_2 = \{y_1, y_2\}$ are binary spaces such that $x_1 \leq x_2$ and $y_1 \leq y_2$, and let $(\underline{F}, \overline{F})$ be a bivariate p -box on $\Omega_1 \times \Omega_2$. Then the following are equivalent:*

1. $(\underline{F}, \overline{F})$ avoids sure loss.
2. $\overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0$ for all $x_1, x_2 \in \Omega_1, y_1, y_2 \in \Omega_2$.
3. $\underline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0$ for all $x_1, x_2 \in \Omega_1, y_1, y_2 \in \Omega_2$.

Proof: The first statement implies the second from Proposition 4.94. To see that the second implies the third note that, since \underline{F} and \overline{F} are standardized maps, it holds that $\underline{F}(x_2, y_2) = \overline{F}(x_2, y_2) = 1$.

To see that the third statement implies the first, let us consider $F : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ given by

$$\begin{aligned} F(x_1, y_1) &= \overline{F}(x_1, y_1) \\ F(x_1, y_2) &= \max\{\overline{F}(x_1, y_1), \underline{F}(x_1, y_2)\} \\ F(x_2, y_1) &= \max\{\overline{F}(x_1, y_1), \underline{F}(x_2, y_1)\} \\ F(x_2, y_2) &= 1. \end{aligned}$$

By construction, F is a standardized map and it is bounded by $\underline{F}, \overline{F}$. To see that it indeed is a distribution function, note that if either $F(x_1, y_2)$ or $F(x_2, y_1)$ is equal to $\overline{F}(x_1, y_1) = F(x_1, y_1)$, then it follows from the monotonicity of $\underline{F}, \overline{F}$ that

$$F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \geq 0;$$

and if $F(x_1, y_2) = \underline{F}(x_1, y_2)$ and $F(x_2, y_1) = \underline{F}(x_2, y_1)$, then

$$\begin{aligned} F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \\ = \underline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0. \quad \blacksquare \end{aligned}$$

Coherence Let us turn now to coherence, where we shall see that Theorem 2.35 does not extend immediately to the bivariate case. We begin by establishing a result related to Proposition 4.93:

Proposition 4.97 *The lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ induced by the bivariate p-box $(\underline{F}, \overline{F})$ is coherent if and only if \underline{F} and \overline{F} are the lower and the upper envelopes of the set*

$$\{F : \Omega_1 \times \Omega_2 \rightarrow [0, 1] \text{ distribution function} : \underline{F} \leq F \leq \overline{F}\},$$

respectively.

Proof: We begin with the direct implication. If $\underline{P}_{(\underline{F}, \overline{F})}$ is coherent, then for any $x \in \Omega_1$ and $y \in \Omega_2$ there is some probability $P \geq \underline{P}_{(\underline{F}, \overline{F})}$ such that $P(A_{(x,y)}) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)})$. Consider the function $F_P : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ defined by $F_P(x', y') = P(A_{(x',y')})$ for every $(x', y') \in \Omega_1 \times \Omega_2$. Reasoning as in the proof of Proposition 4.93, we deduce that F_P is a distribution function that belongs to $(\underline{F}, \overline{F})$. Moreover, by construction:

$$F_P(x, y) = P(A_{(x,y)}) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}) = \underline{F}(x, y).$$

Similarly, there exists some $P' \geq \underline{P}_{(\underline{F}, \overline{F})}$ such that

$$P'(A_{(x,y)}^c) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}^c).$$

Let $F_{P'} : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ be given by $F_{P'}(x', y') = P'(A_{(x',y')})$ for every $(x', y') \in \Omega_1 \times \Omega_2$. Reasoning as in the proof of Proposition 4.93, we deduce that $F_{P'}$ is a distribution function that belongs to $(\underline{F}, \overline{F})$. Moreover, by construction:

$$1 - F_{P'}(x, y) = 1 - P(A_{(x,y)}) = P(A_{(x,y)}^c) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}^c) = 1 - \overline{F}(x, y),$$

whence $F_{P'}(x, y) = \overline{F}(x, y)$.

Conversely, fix $(x, y) \in \Omega_1 \times \Omega_2$ and let F_1, F_2 be distribution functions in $(\underline{F}, \overline{F})$ such that $F_1(x, y) = \underline{F}(x, y)$ and $F_2(x, y) = \overline{F}(x, y)$. Let P_1, P_2 be the finitely additive probabilities they induce in \mathcal{K} by means of Equation (4.15). Then it follows from the proof of Proposition 4.93 that P_1, P_2 dominate $\underline{P}_{(\underline{F}, \overline{F})}$, and moreover

$$\begin{aligned} P_1(A_{(x,y)}) &= F_1(x, y) = \underline{F}(x, y) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}) \text{ and} \\ P_2(A_{(x,y)}^c) &= 1 - P_2(A_{(x,y)}) = 1 - F_2(x, y) = 1 - \overline{F}(x, y) = \underline{P}_{(\underline{F}, \overline{F})}(A_{(x,y)}^c) \end{aligned}$$

Since P_1, P_2 have finitely additive extensions to $\mathcal{P}(\Omega_1 \times \Omega_2)$, we deduce from this that $\underline{P}_{(\underline{F}, \overline{F})}$ is coherent. \blacksquare

We shall call the bivariate p-box $(\underline{F}, \overline{F})$ *coherent* when its associated lower probability is. One interesting difference with the univariate case is that $\underline{F}, \overline{F}$ need not be

distribution functions for $(\underline{F}, \overline{F})$ to be coherent (although if $\underline{F}, \overline{F}$ are distribution functions then trivially $(\underline{F}, \overline{F})$ is coherent by Proposition 4.97). This can be seen for instance with Example 4.90, where the lower envelope of a set of distribution functions (which determines the lower distribution function of a coherent p-box) is not a distribution function itself.

Our next result uses properties (2.11)–(2.15) of coherent lower probabilities to obtain four imprecise-versions of the rectangle inequality that, as we shall see, will play an important role.

Proposition 4.98 *Let $(\underline{F}, \overline{F})$ be a bivariate p-box on $\Omega_1 \times \Omega_2$. If it is coherent, then the following conditions hold for every $x_1, x_2 \in \Omega_1$ and $y_1, y_2 \in \Omega_2$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$:*

$$\begin{aligned} (\mathbf{I} - \mathbf{RI1}) : \quad & \underline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0. \\ (\mathbf{I} - \mathbf{RI2}) : \quad & \overline{F}(x_2, y_2) + \underline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0. \\ (\mathbf{I} - \mathbf{RI3}) : \quad & \overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \overline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0. \\ (\mathbf{I} - \mathbf{RI4}) : \quad & \overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \overline{F}(x_2, y_1) \geq 0. \end{aligned}$$

Proof: Consider (x_1, y_1) and (x_2, y_2) in $\Omega_1 \times \Omega_2$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$. Let $\underline{P}_{(\underline{F}, \overline{F})}$ be the lower probability induced by $(\underline{F}, \overline{F})$ by means of Equation (4.14). It is coherent by Proposition 4.97.

Then, by Equations (2.11) and (2.13), it holds that:

$$\begin{aligned} \underline{P}(A_{(x_2, y_2)}) & \geq \underline{P}(A_{(x_1, y_2)} \cup A_{(x_2, y_1)}) + \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) \\ & \geq \underline{P}(A_{(x_1, y_2)}) + \underline{P}(A_{(x_2, y_1)}) - \overline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}) \\ & \quad + \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})). \end{aligned}$$

Thus:

$$\begin{aligned} \underline{P}(A_{(x_2, y_2)}) - \underline{P}(A_{(x_1, y_2)}) - \underline{P}(A_{(x_2, y_1)}) + \overline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}) \\ \geq \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) \geq 0. \end{aligned}$$

If we write the previous equation in terms of the maps $\underline{F}, \overline{F}$, we obtain that:

$$\underline{F}(x_2, y_2) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) + \overline{F}(x_1, y_1) \geq 0.$$

On the other hand, applying Equations (2.12) and (2.14)

$$\begin{aligned} \overline{P}(A_{(x_2, y_2)}) & \geq \overline{P}(A_{(x_1, y_2)} \cup A_{(x_2, y_1)}) + \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) \\ & \geq \underline{P}(A_{(x_1, y_2)}) + \underline{P}(A_{(x_2, y_1)}) - \underline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}) \\ & \quad + \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})). \end{aligned}$$

Then:

$$\begin{aligned} \overline{P}(A_{(x_2, y_2)}) + \underline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}) - \underline{P}(A_{(x_1, y_2)}) - \underline{P}(A_{(x_2, y_1)}) \\ \geq \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) \geq 0. \end{aligned}$$

In terms of $\underline{F}, \overline{F}$, this means that

$$\overline{F}(x_2, y_2) + \underline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) \geq 0.$$

Analogously, by Equation (2.12)

$$\overline{P}(A_{(x_2, y_2)}) \geq \overline{P}(A_{(x_1, y_2)} \cup A_{(x_2, y_1)}) + \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)}))$$

and, from Equation (2.15), this is greater than or equal to both

$$\underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) + \underline{P}(A_{(x_1, y_2)}) + \overline{P}(A_{(x_2, y_1)}) - \overline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)})$$

and

$$\underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) + \overline{P}(A_{(x_1, y_2)}) + \underline{P}(A_{(x_2, y_1)}) - \overline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}).$$

Then:

$$\begin{aligned} 0 &\leq \underline{P}(A_{(x_2, y_2)} \setminus (A_{(x_1, y_2)} \cup A_{(x_2, y_1)})) \\ &\leq \begin{cases} \overline{P}(A_{(x_2, y_2)}) - \underline{P}(A_{(x_1, y_2)}) - \overline{P}(A_{(x_2, y_1)}) + \overline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}). \\ \overline{P}(A_{(x_2, y_2)}) - \overline{P}(A_{(x_1, y_2)}) - \underline{P}(A_{(x_2, y_1)}) + \overline{P}(A_{(x_1, y_2)} \cap A_{(x_2, y_1)}). \end{cases} \end{aligned}$$

In terms of $\underline{F}, \overline{F}$, this means that:

$$\begin{aligned} \overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \overline{F}(x_1, y_2) - \underline{F}(x_2, y_1) &\geq 0. \\ \overline{F}(x_2, y_2) + \overline{F}(x_1, y_1) - \underline{F}(x_1, y_2) - \overline{F}(x_2, y_1) &\geq 0. \quad \blacksquare \end{aligned}$$

None of these conditions is sufficient for coherence, as we can see in the following examples.

Example 4.99 *Let us show an example where both \underline{F} and \overline{F} satisfy (I-RI1), (I-RI2) and (I-RI4), but not (I-RI3), and the lower prevision \underline{P} is not coherent. For this aim consider three real numbers $a < b < c$ and the functions \underline{F} and \overline{F} defined by:*

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
\underline{F}	0	0.3	0.45	0.3	0.6	0.75	0.45	0.8	1
\overline{F}	0	0.3	0.5	0.3	0.6	0.85	0.5	0.85	1

Both \underline{F} and \overline{F} are standardized maps. In addition, \underline{F} is a distribution function, and consequently \underline{F} and \overline{F} satisfy (I-RI1) and (I-RI2). It can be checked that (I-RI4) is also satisfied. Assume that their lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ is coherent. Then, by Proposition 4.97

there must be a distribution function F between $\underline{F}, \overline{F}$ such that $F(b, c) = \overline{F}(b, c) = 0.85$. However, this implies that

$$F(c, c) + F(b, b) - F(b, c) - F(c, b) = 1 + 0.6 - 0.85 - F(c, b) \geq 0 \Rightarrow F(c, b) \leq 0.75.$$

But on the other hand we must have $F(c, b) \geq \underline{F}(c, b) = 0.8$; this is a contradiction.

Similarly, if we define \underline{F}^* and \overline{F}^* by $\underline{F}^*(x, y) = \underline{F}(y, x)$ and $\overline{F}^*(x, y) = \overline{F}(y, x)$, we obtain an example where (I-RI1), (I-RI2) and (I-RI3) are satisfied but the p-box $(\underline{F}, \overline{F})$ is not coherent. \blacklozenge

Example 4.100 Let us give next an example where \underline{F} and \overline{F} satisfy conditions (I-RI2) and (I-RI3) and (I-RI4), but not (I-RI1), and the bivariate p-box $(\underline{F}, \overline{F})$ is not coherent. For this aim consider three real numbers $a < b < c$ and the functions \underline{F} and \overline{F} defined by:

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
\underline{F}	0	0.3	0.4	0.3	0.6	0.6	0.5	0.8	1
\overline{F}	0	0.3	0.4	0.3	0.6	0.7	0.5	0.8	1

Both \underline{F} and \overline{F} are standardized functions. They also satisfy conditions (I-RI2) and, since \overline{F} is a cumulative distribution function, also conditions (I-RI3) and (I-RI4). Assume that $(\underline{F}, \overline{F})$ is coherent. Then, there must be a distribution function F such that $F(b, c) = \underline{F}(b, c) = 0.6$. Then:

$$F(b, c) + F(a, b) - F(b, b) - F(a, c) = 0.6 + 0.3 - 0.6 - 0.4 = -0.1 < 0,$$

a contradiction. \blacklozenge

Example 4.101 Finally, let us give an example where \underline{F} and \overline{F} satisfy (I-RI1) and (I-RI3) and (I-RI4), but not condition (I-RI2), and the bivariate p-box $(\underline{F}, \overline{F})$ is not coherent. As in the previous examples, consider three real numbers $a < b < c$ and the functions \underline{F} and \overline{F} defined by:

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
\underline{F}	0	0.3	0.4	0.3	0.5	0.7	0.5	0.8	1
\overline{F}	0.1	0.3	0.4	0.3	0.5	0.7	0.5	0.8	1

These functions can be easily proven to satisfy (I-RI1), (I-RI3) and (I-RI4). However, they do not satisfy (I-RI2) since:

$$\overline{F}(b, b) + \underline{F}(a, a) - \underline{F}(a, b) - \underline{F}(b, a) = 0.5 + 0 - 0.3 - 0.3 = -0.1 < 0.$$

Then, $\underline{P}_{(\underline{F}, \overline{F})}$ is not coherent. \blacklozenge

Next we establish the most important result in this section: a characterisation of the coherence of a bivariate p-box in the case when one of the variables is binary.

Proposition 4.102 *Assume that $\Omega_2 = \{y_1, y_2\}$ is a binary space and $\Omega_1 = \{x_1, \dots, x_n\}$ is finite, and let $(\underline{F}, \overline{F})$ be a bivariate p -box on $\Omega_1 \times \Omega_2$.*

1. *If $\underline{F}, \overline{F}$ satisfy (I-RI1) and (I-RI2), then*

$$\underline{F} = \min\{F \text{ distribution function} : \underline{F} \leq F \leq \overline{F}\}.$$

2. *If $\underline{F}, \overline{F}$ satisfy (I-RI3) and (I-RI4), then*

$$\overline{F} = \max\{F \text{ distribution function} : \underline{F} \leq F \leq \overline{F}\}.$$

3. *As a consequence, $(\underline{F}, \overline{F})$ is coherent $\Leftrightarrow \underline{F}, \overline{F}$ satisfy conditions (I-RI1) to (I-RI4).*

Proof: First of all, let us check that if \underline{F} and \overline{F} satisfy (I-RI2), then there is a cumulative distribution function F_2 such that $\underline{F} \leq F_2$ and $F_2(x_i, y_1) = \underline{F}(x_i, y_1)$ for any $i = 1, \dots, n$. For this aim we define the function F_2 by:

$$\begin{aligned} F_2(x_i, y_1) &= \underline{F}(x_i, y_1) \text{ for } i = 1, \dots, n, \\ F_2(x_1, y_2) &= \underline{F}(x_1, y_2), \text{ and} \\ F_2(x_i, y_2) &= \underline{F}(x_i, y_2) - \min(0, \Delta_{\underline{F}}^{R_{i-1}}), \text{ for } i = 2, \dots, n, \text{ where} \\ \Delta_{\underline{F}}^{R_{i-1}} &= \underline{F}(x_i, y_2) + \underline{F}(x_{i-1}, y_1) - \underline{F}(x_i, y_1) - F_2(x_{i-1}, y_2). \end{aligned}$$

On the one hand, by definition $F_2(x_i, y_1) = \underline{F}(x_i, y_1)$ for $i = 1, \dots, n$. On the other hand, let us prove that $\underline{F} \leq F_2 \leq \overline{F}$, $F_2(x_n, y_2) = 1$, F_2 is monotone and $\Delta_{F_2}^{R_{i-1}} \geq 0$, where:

$$\Delta_{F_2}^{R_{i-1}} = F_2(x_i, y_2) + F_2(x_{i-1}, y_1) - F_2(x_i, y_1) - F_2(x_{i-1}, y_2),$$

for $i = 2, \dots, n$. In such a case, F_2 would be a distribution function bounded by \underline{F} and \overline{F} .

1. $F_2 \geq \underline{F}$:

It trivially holds since $-\min(0, \Delta_{\underline{F}}^{R_{i-1}}) \geq 0$.

2. $F_2 \leq \overline{F}$:

For either $i = 1$ or $j = 1$, $F_2(x_i, y_j) = \underline{F}(x_i, y_j) \leq \overline{F}(x_i, y_j)$. When $i, j \geq 2$, and $(i, j) \neq (n, 2)$, it holds that:

$$\overline{F}(x_i, y_2) \geq F_2(x_i, y_2) \Leftrightarrow \overline{F}(x_i, y_2) - \underline{F}(x_i, y_2) + \min(\Delta_{\underline{F}}^{R_{i-1}}, 0) \geq 0$$

This is obvious when $\Delta_{\underline{F}}^{R_{i-1}} \geq 0$. Otherwise, we have to prove that

$$\overline{F}(x_i, y_2) - \underline{F}(x_i, y_2) + \Delta_{\underline{F}}^{R_{i-1}} \geq 0.$$

This inequality holds if and only if:

$$0 \leq \overline{F}(x_i, y_2) - \underline{F}(x_i, y_2) + \underline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - F_2(x_{i-1}, y_2) + \underline{F}(x_{i-1}, y_1) \\ = \overline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - F_2(x_{i-1}, y_2) + \underline{F}(x_{i-1}, y_1).$$

Then, we shall prove that

$$\overline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - F_2(x_k, y_2) + \underline{F}(x_k, y_1) \geq 0 \quad (4.16)$$

for any $k = 1, \dots, i - 1$ by induction on k .

(a) $k = 1$: Equation (4.16) becomes:

$$\overline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - \underline{F}(x_1, y_2) + \underline{F}(x_1, y_1) \geq 0,$$

and it holds for (I-RI2).

(b) Assume that Equation (4.16) holds for $k - 1$. Then, for $k = 1$ Equation (4.16) becomes:

$$\overline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - \underline{F}(x_k, y_2) + \min(\Delta_{\underline{F}}^{R_k-1}, 0) + \underline{F}(x_k, y_1) \geq 0,$$

and this is positive when $\Delta_{\underline{F}}^{R_k-1} \geq 0$ by (I-RI2). Otherwise, it becomes:

$$\overline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - \underline{F}(x_k, y_2) + \underline{F}(x_k, y_2) - \underline{F}(x_k, y_1) \\ - F_2(x_{k-1}, y_2) + \underline{F}(x_{k-1}, y_1) + \underline{F}(x_k, y_1) \\ = \overline{F}(x_i, y_2) - \underline{F}(x_i, y_1) - F_2(x_{k-1}, y_2) + \underline{F}(x_{k-1}, y_1) \geq 0,$$

since Equation (4.16) holds for $k - 1$.

3. $F_2(x_n, y_2) = 1$:

In fact:

$$F_2(x_n, y_2) = 1 \Leftrightarrow \underline{F}(x_n, y_2) - \min(\Delta_{\underline{F}}^{R_n-1}, 0) = 1 - \min(\Delta_{\underline{F}}^{R_n-1}, 0) = 1 \\ \Leftrightarrow \Delta_{\underline{F}}^{R_n-1} \geq 0 \\ \Leftrightarrow \underline{F}(x_n, y_2) - \underline{F}(x_n, y_1) - F_2(x_{n-1}, y_2) + \underline{F}(x_{n-1}, y_1) \\ = \overline{F}(x_n, y_2) - \underline{F}(x_n, y_1) - F_2(x_{n-1}, y_2) + \underline{F}(x_{n-1}, y_1) \geq 0,$$

which follows from the proof by induction of Equation (4.16) by putting $i = n$ and $k = n - 1$.

4. F_2 is monotone:

(a) On the one hand, $F_2(x_i, y_1) = \underline{F}(x_i, y_1) \leq \underline{F}(x_{i+1}, y_1) = F_2(x_i, y_1)$ for any $i = 1, \dots, n - 1$.

(b) $F_2(x_i, y_2) \geq F_2(x_{i-1}, y_2)$:

$$\begin{aligned} F_2(x_i, y_2) &= \underline{F}(x_i, y_2) - \min(\Delta_{\underline{F}}^{R_{i-1}}, 0) \\ &= \max(\underline{F}(x_i, y_2) - \Delta_{\underline{F}}^{R_{i-1}}, \underline{F}(x_i, y_2)) \\ &= \max(F_2(x_{i-1}, y_2) + \underline{F}(x_i, y_1) - \underline{F}(x_{i-1}, y_1), \underline{F}(x_i, y_2)) \\ &\geq F_2(x_{i-1}, y_2) + \underline{F}(x_i, y_1) - \underline{F}(x_{i-1}, y_1) \geq F_2(x_{i-1}, y_2), \end{aligned}$$

by the monotonicity of \underline{F} .

(c) $F_2(x_i, y_2) \geq F_2(x_i, y_1) = \underline{F}(x_i, y_1)$ since

$$F_2(x_i, y_2) \geq \underline{F}(x_i, y_2) \geq \underline{F}(x_i, y_1).$$

5. $\Delta_{F_2}^{R_{i-1}} \geq 0$ for $i = 1, \dots, n$:

It holds that:

$$\begin{aligned} \Delta_{F_2}^{R_{i-1}} &= F_2(x_i, y_2) - \underline{F}(x_i, y_1) - F_2(x_{i-1}, y_2) + \underline{F}(x_{i-1}, y_1) \\ &= \underline{F}(x_i, y_2) + \max(-\Delta_{\underline{F}}^{R_{i-1}}, 0) - \underline{F}(x_i, y_1) - F_2(x_{i-1}, y_2) + \underline{F}(x_{i-1}, y_1) \\ &= \max(-\Delta_{\underline{F}}^{R_{i-1}}, 0) + \Delta_{\underline{F}}^{R_{i-1}} = \max(0, \Delta_{\underline{F}}^{R_{i-1}}) \geq 0. \end{aligned}$$

Now, consider the function F_1 defined by:

$$\begin{aligned} F_1(x_i, y_2) &= \underline{F}(x_i, y_2) \text{ for } i = 1, \dots, n, \\ F_1(x_i, y_1) &= \underline{F}(x_i, y_1) - \min(\Delta_{\underline{F}}^{R_i}, 0), \text{ where} \\ \Delta_{\underline{F}}^{R_i} &= \underline{F}(x_{i+1}, y_2) - F_1(x_{i+1}, y_1) - \underline{F}(x_i, y_2) + \underline{F}(x_i, y_1), \end{aligned}$$

for $i = n-1, \dots, 1$. If \underline{F} and \overline{F} satisfy (I-RI1), with a similar proof as the one for F_2 , we can prove that F_1 is a distribution function bounded by \underline{F} and \overline{F} and, by its definition, $F_1(x_i, y_2) = \underline{F}(x_i, y_2)$ for $i = 1, \dots, n$. Then, taking into account F_1 and F_2 , it holds that:

$$\underline{F} = \min\{F \text{ distribution functions} : \underline{F} \leq F \leq \overline{F}\}.$$

Finally, consider the functions F_3 and F_4 , defined by:

$$\begin{aligned} F_3(x_i, y_2) &= \overline{F}(x_i, y_2) \text{ for } i = 1, \dots, n, \\ F_3(x_1, y_1) &= \overline{F}(x_1, y_1), \text{ and} \\ F_3(x_i, y_1) &= \overline{F}(x_i, y_1) + \min(\Delta_{\overline{F}}^{R_{i-1}}, 0), \text{ where} \\ \Delta_{\overline{F}}^{R_{i-1}} &= \overline{F}(x_i, y_2) + F_3(x_{i-1}, y_1) - \overline{F}(x_{i-1}, y_2) - \overline{F}(x_i, y_1) \end{aligned}$$

for $i = 2, \dots, n$, and:

$$\begin{aligned} F_4(x_i, y_1) &= \overline{F}(x_i, y_1) \text{ for } i = 1, \dots, n, \\ F_4(x_n, y_2) &= \overline{F}(x_n, y_2), \text{ and} \\ F_4(x_i, y_2) &= \overline{F}(x_i, y_2) + \min(\Delta_{\overline{F}}^{R_i}, 0), \text{ where} \\ \Delta_{\overline{F}}^{R_i} &= \overline{F}(x_{i+1}, y_2) + \overline{F}(x_i, y_1) - F_4(x_i, y_2) - \overline{F}(x_{i-1}, y_1) \end{aligned}$$

for $i = n - 1, \dots, 1$. With a similar proof as the one for F_2 , we can check that when \underline{F} and \overline{F} satisfy (I-RI3) (respectively (I-RI4)) F_3 (respectively F_4) is a distribution function bounded by \underline{F} and \overline{F} such that $F_3(x_i, y_2) = \overline{F}(x_i, y_2)$ (respectively $F_4(x_i, y_1) = \overline{F}(x_i, y_1)$) for $i = 1, \dots, n$. Then, this implies that when \underline{F} and \overline{F} satisfy conditions (I-RI3) and (I-RI4) it holds that:

$$\overline{F} = \max\{F \text{ distribution functions} : \underline{F} \leq F \leq \overline{F}\}.$$

Putting the functions F_1, F_2, F_3 and F_4 together, we deduce that when \underline{F} and \overline{F} satisfy (I-RI1) to (I-RI4), $(\underline{F}, \overline{F})$ is a coherent bivariate p-box; the converse implication holds by Proposition 4.98. ■

As a consequence, we deduce that conditions (I-RI1)–(I-RI4) are also equivalent to the coherence of $(\underline{F}, \overline{F})$ when both variables Ω_1, Ω_2 are binary. In fact, we conjecture that conditions (I-RI1)–(I-RI4) are also equivalent to the coherence of $(\underline{F}, \overline{F})$ in the general case.

To conclude this section, we investigate if the third statement in Theorem 2.35 can be used to characterise coherence in the bivariate case. Let $\underline{F}, \overline{F}$ be standardized maps on $\Omega_1 \times \Omega_2$, and let $\underline{P}_{\underline{F}} : \mathcal{K}_1 \rightarrow \mathbb{R}$ and $\underline{P}_{\overline{F}} : \mathcal{K}_2 \rightarrow \mathbb{R}$ be the lower probabilities associated with them by Equation (4.13).

Proposition 4.103 *Let $(\underline{F}, \overline{F})$ be a bivariate p-box and let $\underline{P}_{\underline{F}}, \underline{P}_{\overline{F}}$ be the lower previsions they induce on $\mathcal{K}_1, \mathcal{K}_2$, respectively. Then:*

- (a) $\underline{P}_{\underline{F}}, \underline{P}_{\overline{F}}$ always avoid sure loss.
- (b) $\underline{P}_{\underline{F}}$ is coherent $\Leftrightarrow \underline{P}_{(\underline{F}, 1)}$ is coherent.
- (c) $\underline{P}_{\overline{F}}$ is coherent $\Leftrightarrow \underline{P}_{(I_{(x^*, y^*)}, \overline{F})}$ is coherent.
- (d) $\underline{P}_{(\underline{F}, \overline{F})}$ coherent $\Rightarrow \underline{P}_{\underline{F}}, \underline{P}_{\overline{F}}$ coherent.

Proof:

- (a) To see that $\underline{P}_{\underline{F}}$ and $\underline{P}_{\overline{F}}$ always avoid sure loss, it suffices to take into account that the constant map on 1 is a distribution function that dominates \underline{F} and that $I_{(x^*, y^*)}$ is a distribution function that is dominated by \overline{F} .
- (b) The lower probability $\underline{P}_{\underline{F}}$ is coherent if and only if for every $(x, y) \in \Omega_1 \times \Omega_2$ there is a distribution function $F \geq \underline{F}$ such that $F(x, y) = \underline{F}(x, y)$. The condition $F \geq \underline{F}$ is equivalent to $\underline{F} \leq F \leq 1$, and on the other hand the constant map on 1 is trivially a distribution function. We deduce from Proposition 4.97 that $\underline{P}_{(\underline{F}, 1)}$ is coherent if and only if \underline{F} is the lower envelope of the distribution functions in $(\underline{F}, 1)$, and as a consequence we have the equivalence.

- (c) The lower probability $\underline{P}_{\overline{F}}$ is coherent if and only if for every $(x, y) \in \Omega_1 \times \Omega_2$ there is a distribution function $F \leq \overline{F}$ such that $F(x, y) = \overline{F}(x, y)$. The condition $F \leq \overline{F}$ is equivalent to $I_{(x^*, y^*)} \leq F \leq \overline{F}$, and on the other hand the map $I_{(x^*, y^*)}$ is trivially a distribution function. We deduce from Proposition 4.97 that $\underline{P}_{(I_{(x^*, y^*)}, \overline{F})}$ is coherent if and only if \overline{F} is the upper envelope of the distribution functions in $(I_{(x^*, y^*)}, \overline{F})$, and as a consequence we have the equivalence.
- (d) This statement follows from the previous two and from Proposition 4.97, taking into account that the set of distribution functions $(\underline{F}, \overline{F})$ is the intersection of the sets $(\underline{F}, 1)$ and $(I_{(x^*, y^*)}, \overline{F})$. ■

To see that the converse in the fourth statement does not hold, consider the following example.

Example 4.104 Consider now the functions \underline{F} and \overline{F} of Example 4.100. To see that $(\underline{F}, 1)$ is coherent, it suffices to take into account that \underline{F} is the lower envelope of the distribution functions F_1, F_2 given by:

	(a, a)	(a, b)	(a, c)	(b, a)	(b, b)	(b, c)	(c, a)	(c, b)	(c, c)
F_1	0	0.3	0.4	0.3	0.6	0.7	0.5	0.8	1
F_2	0.1	0.4	0.4	0.3	0.6	0.6	0.5	0.8	1

while the constant map on 1 is trivially a distribution function.

Similarly, since both $I_{(c,c)}$ and \overline{F} are distribution functions, we deduce that $(I_{(c,c)}, \overline{F})$ is also coherent. However, we saw in Example 4.100 that $(\underline{F}, \overline{F})$ are not coherent. ♦

This shows that one of the equivalences in Theorem 2.35 does not extend to the bivariate case. Moreover, we can see from this example that the coherence of $\underline{P}_{\underline{F}}$ does not imply that \underline{F} is a distribution function: we have that $\underline{F}(a, b) + \underline{F}(b, c) < \underline{F}(a, c) + \underline{F}(b, b)$. In a similar way (using for instance Example 4.99) we can see that the coherence of $\underline{P}_{\overline{F}}$ does not imply that \overline{F} is a distribution function.

Another consequence is that whenever (I-RI1)–(I-RI4) characterise the coherence of $(\underline{F}, \overline{F})$ (as is for instance the case in Proposition 4.102), it holds that $\underline{P}_{\overline{F}}$ is coherent for any a standardized function \overline{F} , because they hold trivially whenever \underline{F} is the indicator function $I_{(x^*, y^*)}$. On the other hand, $\underline{P}_{\underline{F}}$ may not be coherent: consider $\Omega_1 = \Omega_2 = \{0, 1\}$ and F given by:

	$(0, 0)$	$(0, 1)$	$(1, 0)$	$(1, 1)$
\underline{F}	0	0.6	0.6	1

Then there is no distribution function $F \geq \underline{F}$ satisfying $F(0, 0) = \underline{F}(0, 0) = 0$, because then

$$F(1, 1) + F(0, 0) = 1 < 1.2 \leq F(0, 1) + F(1, 0).$$

2-monotonicity In the univariate case, the lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ associated with a p-box is completely monotone [198]. As we saw in Definition 2.40, this means, in particular, that for every pair of events A, B in its domain it holds that

$$\underline{P}_{(\underline{F}, \overline{F})}(A \cup B) + \underline{P}_{(\underline{F}, \overline{F})}(A \cap B) \geq \underline{P}_{(\underline{F}, \overline{F})}(A) + \underline{P}_{(\underline{F}, \overline{F})}(B),$$

provided also $A \cup B$ and $A \cap B$ belong to the domain. 2-monotone capacities have been studied in detail in [53, 204], among others. They satisfy the property of *comonotone additivity*, which is of interest in economy ([35, 203]).

In the univariate case, we can assume without loss of generality that the domain of the lower probability induced by the p-box is a lattice (see [198] for more details), and this allows us to apply the results from [53]. This is not the case for bivariate p-boxes: the domain \mathcal{K} of $\underline{P}_{(\underline{F}, \overline{F})}$ is not a lattice, so if we want to use the results in [53] we need to take the natural extension of $\underline{P}_{(\underline{F}, \overline{F})}$. By the Envelope Theorem (Theorem 2.30) and Proposition 4.97, this natural extension is the lower envelope of the set

$$\{P_F : F \text{ distribution function, } \underline{F} \leq F \leq \overline{F}\},$$

where P_F is the finitely additive probability associated with the distribution function F by means of Equation (4.15).

However, and as the following example shows, in the bivariate case it could be that the lower probability associated with the p-box $(\underline{F}, \overline{F})$ is coherent but not 2-monotone, even if both $\underline{F}, \overline{F}$ are distribution functions:

Example 4.105 Consider $\Omega_1 = \Omega_2 = \{0, 1\}$, and let $\underline{F}, \overline{F} : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ be the standardized maps given by:

	(0, 0)	(0, 1)	(1, 0)	(1, 1)
\underline{F}	0	0	0.5	1
\overline{F}	0.25	0.25	0.5	1

Then, both $\underline{F}, \overline{F}$ are distribution functions, because

$$\begin{aligned} \underline{F}(1, 1) + \underline{F}(0, 0) - \underline{F}(0, 1) - \underline{F}(1, 0) &= 0; \\ \overline{F}(1, 1) + \overline{F}(0, 0) - \overline{F}(0, 1) - \overline{F}(1, 0) &= 0.25 > 0; \end{aligned}$$

and the other comparisons are trivial.

Now, in the particular case of binary spaces the correspondence between distribution functions and finitely additive probabilities in Equation (4.15) means that any distribution function F on $\Omega_1 \times \Omega_2$ determines uniquely a probability mass function on $\mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2)$ by:

$$\begin{aligned} P_F(\{(0, 0)\}) &= F(0, 0). \\ P_F(\{(0, 1)\}) &= F(0, 1) - F(0, 0). \\ P_F(\{(1, 0)\}) &= F(1, 0) - F(0, 0). \\ P_F(\{(1, 1)\}) &= 1 - P_F(\{(0, 1)\}) - P_F(\{(1, 0)\}) - P_F(\{(0, 0)\}) \\ &= F(1, 1) - F(0, 1) - F(1, 0) + F(0, 0) \geq 0. \end{aligned}$$

Let \mathcal{F} be the set of distribution functions that lie between \underline{F} and \overline{F} , and let us define

$$\mathcal{M}_{\mathcal{F}} := \{P_F : F \in \mathcal{F}\}.$$

Then $\underline{P}_{(\underline{F}, \overline{F})}$ is the lower envelope of $\mathcal{M}_{\mathcal{F}}$ on \mathcal{K} and so is its natural extension \underline{E} . Let us show that \underline{E} is not 2-monotone.

Since $\overline{F}(1, 0) = 0.5$, $\overline{F}(0, 1) = 0.25$ and $\underline{F}(1, 1) = 1$, any map F bounded between \underline{F} and \overline{F} will satisfy $F(1, 0) + F(0, 1) \leq F(0, 0) + F(1, 1)$, so it will be a distribution function as soon as it is monotone. In other words, $\mathcal{F} = \{F \text{ monotone} : \underline{F} \leq F \leq \overline{F}\}$.

Denote $a = \{(0, 0)\}$, $b = \{(0, 1)\}$, $c = \{(1, 0)\}$, $d = \{(1, 1)\}$ and take $A = \{a, c\}$ and $B = \{c, d\}$. Any monotone map F bounded by $\underline{F}, \overline{F}$ induces the mass function $(P(a), P(b), P(c), P(d))$, where:

$$\begin{aligned} P(a) &\in [0, 0.25], & P(a) + P(b) &\in [0, 0.25], \\ P(a) + P(c) &= 0.5, & P(a) + P(b) + P(c) + P(d) &= 1. \end{aligned}$$

Then:

$$\begin{aligned} \mathcal{M}_{\mathcal{F}} &= \{(P_F(a), P_F(b), P_F(c), P_F(d)) : F \in (\underline{F}, \overline{F})\} \\ &= \{(\lambda, \nu - \lambda, 0.5 - \lambda, 0.5 - \nu + \lambda) : \nu \in [0, 0.25], \lambda \in [0, \nu]\}, \end{aligned}$$

and as a consequence:

- $\underline{E}(A) = \underline{E}(\{a, c\}) = 0.5$.
- $\underline{E}(B) = \min\{P(c) + P(d) : P \in \mathcal{M}_{\mathcal{F}}\} = 0.75$, considering the mass function $P = (0.25, 0, 0.25, 0.5)$.
- $\underline{E}(A \cup B) = \min\{P(a) + P(c) + P(d) : P \in \mathcal{M}_{\mathcal{F}}\} = 0.75$, with $P = (0, 0.25, 0.5, 0.25)$.
- $\underline{E}(A \cap B) = \min\{P(c) : P \in \mathcal{M}_{\mathcal{F}}\} = 0.25$, considering the mass function $P = (0.25, 0, 0.25, 0.5)$.

This means that $\underline{E}(A \cup B) + \underline{E}(A \cap B) < \underline{E}(A) + \underline{E}(B)$ and therefore the lower probability induced by the p-box $(\underline{F}, \overline{F})$ is not 2-monotone. ♦

Interestingly, in this example the lower probability \underline{E} does not coincide with the lower envelope of $\min\{P_{\underline{F}}, P_{\overline{F}}\}$: these are associated with the mass functions $P_{\underline{F}} = (0, 0, 0.5, 0.5)$ and $P_{\overline{F}} = (0.25, 0, 0.25, 0.5)$, so

$$\min\{P_{\underline{F}}(A \cup B), P_{\overline{F}}(A \cup B)\} = 1 > 0.75 = \underline{E}(A \cup B).$$

This means that even if the p-box is determined by the distribution functions $\underline{F}, \overline{F}$, the same does not apply to its associated lower probability.

On the other hand, when the bivariate p-box determines a 2-monotone lower probability, it is not too difficult to show that \underline{F} is indeed a distribution function. Note here the difference with the case where we only require that the lower probability is coherent, discussed in Section 4.3.1.

Proposition 4.106 ([185, Lemma 6]) *Assume that the natural extension of the lower probability $\underline{P}_{(\underline{F}, \overline{F})}$ induced by the bivariate p-box $(\underline{F}, \overline{F})$ by Equation (4.14) is 2-monotone. Then \underline{F} is a distribution function.*

However, the standardized map \overline{F} of the p-box determined by a 2-monotone lower probability is not necessarily a distribution function.

Example 4.107 *Consider the upper probability defined by $\overline{P}(A) = \min((1 + \delta)P(A), 1)$ for every $A \in \mathcal{P}(\Omega_1 \times \Omega_2)$, where $\delta > 0$,*

$$\mathcal{K} \supseteq \{A_{(x,y)} : x \in \Omega_1, y \in \Omega_2\},$$

and P is a probability measure. This corresponds to Pari-mutuel model (see [205, Section 2.9.3]) and it is known that \overline{P} is 2-alternating. Consider the random variables X and Y defined on $\Omega_1 = \Omega_2 = \{a, b, c\}$, where $a < b < c$, probability P and value of $\delta = 0.25$:

$X \backslash Y$	a	b	c	$X \backslash Y$	a	b	c
a	0.1	0	0.15	a	0.1	0.1	0.25
b	0.2	0.2	0.05	b	0.3	0.5	0.7
c	0.15	0.1	0.05	c	0.45	0.75	1
Joint probability distribution				Joint distribution function			

In this situation, \overline{F} is not a precise cumulative distribution function:

$$\overline{F}(3, 3) + \overline{F}(2, 2) - \overline{F}(3, 2) - \overline{F}(2, 3) = 1 + 0.625 - 0.9375 - 0.875 < 0. \blacklozenge$$

Remark 4.108 *One interesting case is that when the bivariate p-box is precise, that is, when the standardized maps $\underline{F}, \overline{F}$ coincide. In that case, we obviously have that $(\underline{F}, \overline{F})$ avoids sure loss if and only if it is coherent, and if and only if $\underline{F} = \overline{F}$ is a bivariate distribution function. When Ω_1 and Ω_2 are finite, it follows from Equation (4.15) that this distribution function has a unique extension to the power set of $\Omega_1 \times \Omega_2$; this means that in that case the lower probability associated with $(\underline{F}, \overline{F})$ is linear.*

Note however, that a distribution function does not determine uniquely its associated finitely additive probability, not even in the univariate case; this is a problem that has been explored in detail in [133]. \blacklozenge

4.3.2 Imprecise copulas

One particular case where bivariate p-boxes can arise is in the combination of two marginal p-boxes. In this section, we shall explore this case in detail, by studying the properties of a number of bivariate p-boxes with given marginals: the most conservative one, that shall be obtained by means of the Fréchet bounds and the notion of natural extension, and also the one corresponding to model a notion of independence. In both cases, we shall see that the bivariate model can be derived by means of an appropriate extension of the notion of copula.

Related results can be found in [198, Section 7], with one fundamental difference: in [198], the authors assume the existence of a total preorder on the product space $\Omega_1 \times \Omega_2$ that is compatible with the orders in Ω_1, Ω_2 ; while here we shall only consider the partial order given by

$$(x_1, y_1) \leq (x_2, y_2) \Leftrightarrow x_1 \leq x_2 \text{ and } y_1 \leq y_2.$$

An imprecise version of Sklar's theorem

Taking into account our previous results, we see that the combination of the marginal p-boxes into a bivariate one is related to the combination of marginal lower probabilities into a joint one. This is a problem that has been studied in detail under some conditions of independence [52].

Remember that Sklar's Theorem (see Theorem 2.27) stated that given two random variables X and Y with associated cumulative distribution functions F_X and F_Y , there exists a copula C such that the joint distribution function, named F , can be expressed by:

$$F(x, y) = C(F_X(x), F_Y(y)) \text{ for any } x, y.$$

Moreover, the copula is unique on $\text{Rang}(F_X) \times \text{Rang}(F_Y)$. Conversely, any transformation of marginal distribution functions by means of a copula produces a bivariate distribution function.

Next, we introduce the notion of imprecise copula. It is a simple generalisation of precise copulas; the main difference lies in the rectangle inequality that has been replaced by its four imprecise extensions of (I-RI1)–(I-RI4).

Definition 4.109 *A pair of functions $\underline{C}, \overline{C} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is called an imprecise copula if:*

- Both \underline{C} and \overline{C} are component-wise increasing.
- $\underline{C} \leq \overline{C}$.

- $\underline{C}(0, u) = \overline{C}(0, u) = 0 = \overline{C}(v, 0) = \underline{C}(v, 0) \forall u, v \in [0, 1]$.
- $\underline{C}(1, u) = \overline{C}(1, u) = u$ and $\overline{C}(v, 1) = \underline{C}(v, 1) = v \forall u \in S_2, v \in S_1$.
- \underline{C} and \overline{C} satisfy the following conditions for any $x_1, x_2, y_1, y_2 \in [0, 1]$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$:

$$\text{(I - CRI1)} : \overline{C}(x_1, y_1) + \underline{C}(x_2, y_2) \geq \underline{C}(x_1, y_2) + \overline{C}(x_2, y_1).$$

$$\text{(I - CRI2)} : \underline{C}(x_1, y_1) + \overline{C}(x_2, y_2) \geq \overline{C}(x_1, y_2) + \underline{C}(x_2, y_1).$$

$$\text{(I - CRI3)} : \overline{C}(x_1, y_1) + \overline{C}(x_2, y_2) \geq \underline{C}(x_1, y_2) + \overline{C}(x_2, y_1).$$

$$\text{(I - CRI4)} : \overline{C}(x_1, y_1) + \overline{C}(x_2, y_2) \geq \overline{C}(x_1, y_2) + \underline{C}(x_2, y_1).$$

\underline{C} and \overline{C} shall be named the lower and the upper copulas, respectively.

Note that monotonicity and condition $\underline{C} \leq \overline{C}$ may not be imposed in the definition of imprecise copula: on the one hand, $\underline{C} \leq \overline{C}$ can be derived from conditions (I-CRI1) to (I-CRI4): for any $x, y \in [0, 1]$, (I-CRI1) assures that

$$\overline{C}(x, y) + \underline{C}(x, y) \geq \underline{C}(x, y) + \underline{C}(x, y),$$

that is equivalent to $\overline{C}(x, y) \geq \underline{C}(x, y)$. Furthermore, taking $0 \leq x$ and $y_1 \leq y_2$ and applying (I-CRI1) we obtain that \underline{C} is increasing in the second component. Similarly, using conditions (I-CRI1) to (I-CRI4) we obtain that both \underline{C} and \overline{C} are increasing in each component.

As next result shows, one way of obtaining imprecise copulas is by taking the infimum and supremum of sets of copulas, or just simply by considering two ordered copulas.

Proposition 4.110 *Let \mathcal{C} be a non-empty set of copulas. Take \underline{C} and \overline{C} defined by:*

$$\underline{C}(x, y) = \inf_{C \in \mathcal{C}} C(x, y) \text{ and } \overline{C}(x, y) = \sup_{C \in \mathcal{C}} C(x, y)$$

for any (x, y) . Then, $(\underline{C}, \overline{C})$ forms an imprecise copula. Moreover, if C_1 and C_2 are two copulas such that $C_1 \leq C_2$, then (C_1, C_2) also forms an imprecise copula.

Proof: Consider \mathcal{C} a non-empty set of copulas, and let \underline{C} and \overline{C} denote their infimum and supremum. Since any copula is in particular a bivariate cumulative distribution function, $(\underline{C}, \overline{C})$ forms a bivariate p-box. Hence, \underline{C} and \overline{C} satisfy $\underline{C} \leq \overline{C}$, monotonicity, the boundary conditions and (I-CRI1) to (I-CRI4).

In particular, if we consider two copulas C_1 and C_2 such that $C_1 \leq C_2$, the previous result applies, being C_1 and C_2 the infimum and supremum, respectively. ■

Let us see to which extent Sklar's theorem also holds in an imprecise framework. For this aim, we start by considering marginal imprecise distributions, described by (univariate) p-boxes, and we use imprecise copulas to obtain a bivariate p-box that generates a coherent lower probability.

Proposition 4.111 *Let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ be two marginal p-boxes on respective spaces Ω_1, Ω_2 , and let \mathcal{C} be a set of copulas. Define the bivariate p-box $(\underline{F}, \overline{F})$ by:*

$$\underline{F}(x, y) = \inf_{C \in \mathcal{C}} C(\underline{F}_X(x), \underline{F}_Y(y)) \text{ and } \overline{F}(x, y) = \sup_{C \in \mathcal{C}} C(\overline{F}_X(x), \overline{F}_Y(y)) \quad (4.17)$$

for any (x, y) , and let \underline{P} be its associated lower probability by Equation (4.14). Then, \underline{P} is a coherent lower probability. Moreover,

$$\underline{F}(x, y) = \underline{C}(\underline{F}_X(x), \underline{F}_Y(y)) \text{ and } \overline{F}(x, y) = \overline{C}(\overline{F}_X(x), \overline{F}_Y(y)),$$

where $\underline{C}(x, y) = \inf_{C \in \mathcal{C}} C(x, y)$ and $\overline{C}(x, y) = \sup_{C \in \mathcal{C}} C(x, y)$.

Proof: Given $C \in \mathcal{C}$, $F_1 \in (\underline{F}_X, \overline{F}_X)$ and $F_2 \in (\underline{F}_Y, \overline{F}_Y)$, the bivariate distribution function $C(F_1, F_2)$ is bounded by $\underline{F}, \overline{F}$. Applying Proposition 4.93, we deduce that \underline{P} avoids sure loss. Let us now check that it is also coherent. Fix (x, y) in $\Omega_1 \times \Omega_2$. Since the marginal p-boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$ are coherent, there are $F_1 \in (\underline{F}_X, \overline{F}_X)$ and $F_2 \in (\underline{F}_Y, \overline{F}_Y)$ such that $F_1(x) = \underline{F}_X(x)$ and $F_2(y) = \underline{F}_Y(y)$. As a consequence,

$$\underline{F}(x, y) = \inf_{C \in \mathcal{C}} C(\underline{F}_X(x), \underline{F}_Y(y)) = \inf_{C \in \mathcal{C}} C(F_1(x), F_2(y)),$$

and since $C(F_1, F_2) \in (\underline{F}, \overline{F})$ for every $C \in \mathcal{C}$, it then follows from monotonicity that \underline{F} is the lower envelope of the set $\{F \text{ distribution function} : \underline{F} \leq F \leq \overline{F}\}$. Similarly, we can also prove that

$$\overline{F} = \sup\{F \text{ distribution function} : \underline{F} \leq F \leq \overline{F}\}.$$

Applying now Proposition 4.97, we deduce that \underline{P} is coherent. ■

In particular, when the information about the marginal distribution is precise, and it is given by the distribution functions F_X and F_Y , the bivariate p-box in the above proposition is given by

$$\underline{F}(x, y) = \inf_{C \in \mathcal{C}} C(F_X(x), F_Y(y)) \text{ and } \overline{F}(x, y) = \sup_{C \in \mathcal{C}} C(F_X(x), F_Y(y))$$

for any $(x, y) \in \Omega_1 \times \Omega_2$.

Remark 4.112 *This result generalises [167, Theorem 2.4], where the authors only focused on the functions \underline{F} and \overline{F} , showing that*

$$\underline{F}(x, y) = \underline{C}(F_X(x), F_Y(y)) \text{ and } \overline{F}(x, y) = \overline{C}(F_X(x), F_Y(y)).$$

Proposition 4.111 establishes moreover the coherence of the joint lower probability, and it is more general than [167, Theorem 2.4] since we are assuming the existence of imprecision in the marginal distribution, that we model by means of p-boxes. ♦

Using these results, we can give the form of the credal set $\mathcal{M}(\underline{P})$ (that is, the set of dominating probabilities) associated with the lower probability \underline{P} . Note that, in the sequel, we can assume that the probabilities in $\mathcal{M}(\underline{P})$ are defined on a suitable set of events, larger than the domain of \underline{P} . Hence, the domains of \underline{P} and of the probabilities in $\mathcal{M}(\underline{P})$ do not necessarily coincide.

Corollary 4.113 *Under the assumptions of Proposition 4.111, the credal set $\mathcal{M}(\underline{P})$ of the lower probability \underline{P} is given by:*

$$\{P \text{ probability} \mid \underline{C}(\underline{F}_X(x), \underline{F}_Y(y)) \leq F_P(x, y) \leq \overline{C}(\overline{F}_X(x), \overline{F}_Y(y)) \forall x, y\}.$$

Proof: By Proposition 4.97, we know that \underline{P} is coherent if and only if \underline{F} and \overline{F} are the lower and the upper envelopes of the set

$$\{F \text{ distribution function} \mid \underline{F} \leq F \leq \overline{F}\}.$$

From this, the thesis follows simply by replacing the lower and upper distribution functions by their expressions in terms of \underline{C} and \overline{C} . ■

Next, we investigate whether the second part of Sklar's theorem also holds, meaning whether any bivariate p-box can be obtained as the combination of its marginals by means of an imprecise copula. A partial result in this sense has been established in [185, Theorem 9]. The next example shows that this result cannot be generalised to arbitrary p-boxes.

Example 4.114 *Consider $\Omega_1 = \{x_1, x_2, x_3\}, \Omega_2 = \{y_1, y_2\}$ with $x_1 < x_2 < x_3, y_1 < y_2$ and let P_1, P_2 be the probability measures associated with the mass functions:*

	(x_1, y_1)	(x_2, y_1)	(x_1, y_2)	(x_2, y_2)	(x_3, y_1)	(x_3, y_2)
P_1	0.2	0	0.3	0	0	0.5
P_2	0.1	0.2	0.5	0.1	0	0.1

Let $\underline{P} = \min\{P_1, P_2\}$. Then its associated p-box satisfies $\underline{F}(x_1) = \underline{F}(x_2) = 0.5$ and $\underline{F}(y_1) = 0.2$ while $\underline{F}(x_1, y_1) = 0.1 < \underline{F}(x_2, y_1) = 0.2$. Hence, there is no function C such that $\underline{F}(x_1, y_1) = C(\underline{F}(x_1), \underline{F}(y_1)) = C(\underline{F}(x_2), \underline{F}(y_1)) = \underline{F}(x_2, y_1)$. Consequently, the lower distribution in the bivariate p-box cannot be expressed as a function of its marginals. ♦

Obviously, when both $\underline{F}, \overline{F}$ are bivariate distribution functions, we can express them as a function of their marginals because of Sklar's theorem; the example shows that this is no longer possible when they are simply standardized functions.

Next theorem summarises the results of this paragraph.

Theorem 4.115 (Imprecise version of Sklar's Theorem) *Consider a set of copulas \mathcal{C} and two marginal p-boxes $(\underline{F}_X, \overline{F}_X)$. The functions \underline{F} and \overline{F} defined by*

$$\begin{aligned}\underline{F}(x, y) &= \inf_{C \in \mathcal{C}} C(\underline{F}_X(x), \underline{F}_Y(y)) \text{ and} \\ \overline{F}(x, y) &= \sup_{C \in \mathcal{C}} C(\overline{F}_X(x), \overline{F}_Y(y))\end{aligned}$$

form a bivariate p-box whose marginals are $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$. Furthermore, the lower probability associated with this bivariate p-box is coherent.

However, given a bivariate p-box $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$, there may not be an imprecise copula $(\underline{C}, \overline{C})$ that generates $(\underline{F}, \overline{F})$ from its marginals, even when its associated lower probability is coherent.

Natural extension and independent products

In this section we consider two particular combinations of the marginal p-boxes into the bivariate one. First of all, we consider the case where there is no information about the copula that links the marginal distribution functions.

Lemma 4.116 *Consider the univariate p-boxes $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$, and let \underline{P} be the lower prevision defined on*

$$\mathcal{A}^* = \{A_{(x, y^*)}, A_{(x, y^*)}^c, A_{(x^*, y)}, A_{(x^*, y)}^c : x, y \in \mathbb{R}\} \quad (4.18)$$

by

$$\begin{aligned}\underline{P}(A_{(x, y^*)}) &= \underline{F}_X(x) & \underline{P}(A_{(x, y^*)}^c) &= 1 - \overline{F}_X(x). \\ \underline{P}(A_{(x^*, y)}) &= \underline{F}_Y(y) & \underline{P}(A_{(x^*, y)}^c) &= 1 - \overline{F}_Y(y).\end{aligned} \quad (4.19)$$

Then:

1. \underline{P} is a coherent lower probability.
2. $\mathcal{M}(\underline{P}) = \mathcal{M}(C_L, C_M)$, where $\mathcal{M}(C_L, C_M)$ is given by

$$\{P \text{ prob.} \mid F_P(x, y) \in [C_L(\underline{F}_X(x), \underline{F}_Y(y)), C_M(\overline{F}_X(x), \overline{F}_Y(y))]\}.$$

Proof: Let C_P denote the product copula, and let \underline{P}_{C_P} be the coherent lower probability on \mathcal{K} that results from Proposition 4.111, taking $\mathcal{C} = \{C_P\}$. Then \underline{P} coincides with \underline{P}_{C_P} in \mathcal{A}^* , and consequently \underline{P} is coherent.

On the other hand, let us check the equality between the credal sets $\mathcal{M}(\underline{P})$ and $\mathcal{M}(C_L, C_M)$ (note that both sets are trivially non-empty).

- Let P be a probability in $\mathcal{M}(C_L, C_M)$, and let F_P be its associated distribution function. Then it holds that:

$$\begin{aligned} F_P(x, y^*) &\in [C_L(\underline{F}_X(x), 1), C_M(\overline{F}_X(x), 1)] = [\underline{F}_X(x), \overline{F}_X(x)]. \\ F_P(x^*, y) &\in [C_L(1, \underline{F}_Y(y)), C_M(1, \overline{F}_Y(y))] = [\underline{F}_Y(y), \overline{F}_Y(y)]. \end{aligned}$$

Thus, the marginal distribution functions of F_P belong to the p-boxes $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$. As a consequence, $P \in \mathcal{M}(\underline{P})$.

- Conversely, let P be a probability on $\mathcal{M}(\underline{P})$, and let F_P be its associated distribution function. Then, Sklar's Theorem assures that there is a (precise) copula C such that $F_P(x, y) = C(F_P(x, y^*), F_P(x^*, y))$ for every $(x, y) \in \Omega_1 \times \Omega_2$. Hence,

$$\begin{aligned} C_L(\underline{F}_X(x), \underline{F}_Y(y)) &\leq C_L(F_P(x, y^*), F_P(x^*, y)) \leq C(F_P(x, y^*), F_P(x^*, y)) \\ &\leq C(\overline{F}_X(x), \overline{F}_Y(y)) \leq C_M(\overline{F}_X(x), \overline{F}_Y(y)), \end{aligned}$$

taking into account that any copula lies between C_L and C_M . We conclude that $P \in \mathcal{M}(C_L, C_M)$ and as a consequence both sets coincide. ■

From this result we can immediately derive the expression of the *natural extension* [205] of two marginal p-boxes, that is the least-committal (i.e., the most imprecise) coherent lower probability that extends \underline{P} to a larger domain.

Proposition 4.117 *Let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ be two univariate p-boxes. Let \underline{P} be the lower prevision defined on the set \mathcal{A}^* given by Equation (4.18) by means of Equation (4.19). Then, the natural extension \underline{E} of \underline{P} to \mathcal{K} is given by*

$$\underline{E}(A_{(x,y)}) = C_L(\underline{F}_X(x), \underline{F}_Y(y)) \text{ and } \underline{E}(A_{(x,y)}^c) = 1 - C_M(\overline{F}_X(x), \overline{F}_Y(y)).$$

The bivariate p-box $(\underline{F}, \overline{F})$ associated with \underline{E} is given by:

$$\underline{F}(x, y) = C_L(\underline{F}_X(x), \underline{F}_Y(y)) \text{ and } \overline{F}(x, y) = C_M(\overline{F}_X(x), \overline{F}_Y(y)).$$

Proof: On the one hand, the lower prevision \underline{P} is coherent from the previous lemma, and in addition its associated credal set is $\mathcal{M}(\underline{P}) = \mathcal{M}(C_L, C_M)$. The natural extension of \underline{P} to the set \mathcal{K} is given by:

$$\begin{aligned} \underline{E}(A_{(x,y)}) &= \inf_{P \in \mathcal{M}(\underline{P})} F_P(x, y) = \inf_{P \in \mathcal{M}(C_L, C_M)} F_P(x, y) = C_L(\underline{F}_X(x), \underline{F}_Y(y)). \\ \underline{E}(A_{(x,y)}^c) &= \inf_{P \in \mathcal{M}(\underline{P})} (1 - P(A_{(x,y)})) = 1 - \sup_{P \in \mathcal{M}(\underline{P})} F_P(x, y) \\ &= 1 - \sup_{P \in \mathcal{M}(C_L, C_M)} F_P(x, y) = 1 - C_M(\overline{F}_X(x), \overline{F}_Y(y)). \end{aligned}$$

The second part is an immediate consequence of the first. ■

Recall that Proposition 4.110 assures that every pair of copulas C_1 and C_2 satisfying $C_1 \leq C_2$ (in particular C_L and C_M) forms an imprecise copula (C_1, C_2) .

Until now, we have studied how to build the joint p -box $(\underline{F}, \overline{F})$ from two given marginals $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$, when we have no information about the interaction between the underlying variables X and Y : we have argued that we should use in that case the natural extension of the associated coherent lower probabilities, which corresponds to combining the compatible univariate distribution functions by means of all the possible copulas, and then considering the lower envelope.

Next, we consider another case of interest: that where the variables X and Y are assumed to be independent. Consider marginal p -boxes $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$, and let $\underline{P}_X, \underline{P}_Y$ the coherent lower probabilities they induce by means of Equation (2.17). We shall also use this notation to refer to their natural extensions, so that

$$\begin{aligned} \underline{P}_X &:= \min\{P : P(A_x) \in [\underline{F}_X(x), \overline{F}_X(x)] \forall x \in \Omega_1\} \text{ and} \\ \underline{P}_Y &:= \min\{P : P(A_y) \in [\underline{F}_Y(y), \overline{F}_Y(y)] \forall y \in \Omega_2\}. \end{aligned}$$

Under imprecise information, there is more than one way to model the notion of independence; see [47] for a survey on this topic. Because of this, there is more than one manner in which we can say that a coherent lower prevision \underline{P} on the product space is an independent product of its marginals $\underline{P}_X, \underline{P}_Y$. This was studied in some detail in [52]. In the remainder of this paragraph, we shall follow that paper into assuming that the spaces Ω and Ω' are finite. We recall thus the following definitions.

Definition 4.118 *Let \underline{P} be a coherent lower prevision on $\mathcal{L}(\Omega_1 \times \Omega_2)$ with marginals $\underline{P}_X, \underline{P}_Y$. We say that \underline{P} is an independent product when it is coherent with the conditional lower previsions $\underline{P}_X(\cdot|\Omega_2), \underline{P}_Y(\cdot|\Omega_1)$ derived from $\underline{P}_X, \underline{P}_Y$ by epistemic irrelevance, meaning that*

$$\underline{P}_X(f|y) := \underline{P}_X(f(\cdot, y)) \text{ and } \underline{P}_Y(f|x) := \underline{P}_Y(f(x, \cdot)) \forall f \in \mathcal{L}(\Omega_1 \times \Omega_2), x \in \Omega_1, y \in \Omega_2.$$

One example of independent product is the *strong product*, given by

$$\underline{P}_X \boxtimes \underline{P}_Y := \inf\{P_X \times P_Y : P_X \geq \underline{P}_X, P_Y \geq \underline{P}_Y\}.$$

This is the joint model satisfying the notion of *strong independence*. However, it is not the only independent product, nor is it the smallest one. In fact, the smallest independent product of the marginal coherent lower previsions $\underline{P}_X, \underline{P}_Y$ is their *independent natural extension*, which is given by

$$\begin{aligned} (\underline{P}_X \otimes \underline{P}_Y)(f) \\ = \sup\{\mu : f - \mu \geq g - \underline{P}_X(g|\Omega_2) + h - \underline{P}_Y(h|\Omega_1) \text{ for some } g, h \in \mathcal{L}(\Omega_1 \times \Omega_2)\} \end{aligned}$$

for every gamble f on $\Omega_1 \times \Omega_2$.

One way of building independent products is by means of the following condition:

Definition 4.119 *A coherent lower prevision \underline{P} on $\mathcal{L}(\Omega_1 \times \Omega_2)$ is called factorising when*

$$\underline{P}(fg) = \underline{P}(f\underline{P}(g)) \quad \forall f \in \mathcal{L}^+(\Omega_1), g \in \mathcal{L}(\Omega_2)$$

and

$$\underline{P}(fg) = \underline{P}(g\underline{P}(f)) \quad \forall f \in \mathcal{L}(\Omega_1), g \in \mathcal{L}^+(\Omega_2).$$

Both the independent natural extension and the strong product are factorising. Indeed, it can be proven [52, Theorem 28] that any factorising \underline{P} is an independent product of its marginals, but the converse is not true. Under factorisation, it is not difficult to establish the following result.

Proposition 4.120 *Let $(\underline{F}_X, \overline{F}_X), (\underline{F}_Y, \overline{F}_Y)$ be marginal p -boxes, and let $\underline{P}_X, \underline{P}_Y$ be their associated coherent lower previsions. Let \underline{P} be a factorising coherent lower prevision on $\mathcal{L}(\Omega_1 \times \Omega_2)$ with these marginals. Then it induces the bivariate p -box $(\underline{F}, \overline{F})$ given by*

$$\underline{F}(x, y) = \underline{F}_X(x) \cdot \underline{F}_Y(y) \quad \text{and} \quad \overline{F}(x, y) = \overline{F}_X(x) \cdot \overline{F}_Y(y).$$

Proof: It suffices to take into account that, if \underline{P} is factorising, then

$$\underline{P}(A_{(x,y)}) = \underline{P}(I_{A_{(x,y^*)}} \cdot I_{A_{(x^*,y)}}) = \underline{P}(A_{(x,y^*)}) \cdot \underline{P}(A_{(x^*,y)}) = \underline{F}_X(x) \cdot \underline{F}_Y(y),$$

and similarly using conjugacy we deduce that

$$\overline{P}(A_{(x,y)}) = \overline{P}(A_{(x,y^*)} \cdot A_{(x^*,y)}) = \overline{P}(A_{(x,y^*)}) \cdot \overline{P}(A_{(x^*,y)}) = \overline{F}_X(x) \cdot \overline{F}_Y(y),$$

taking into account in the application of the factorisation condition that both gambles $A_{(x,y^*)}, A_{(x^*,y)}$ are positive, and recalling also that x^*, y^* denote the maxima of Ω, Ω' , respectively. ■

From this, it is easy to deduce that the p -box $(\underline{F}, \overline{F})$ induced by a factorising \underline{P} is the lower envelope of the set of bivariate distribution functions

$$\{F : F(x, y) = F_X(x) \cdot F_Y(y) \text{ for } F_X \in (\underline{F}_X, \overline{F}_X), F_Y \in (\underline{F}_Y, \overline{F}_Y)\}.$$

In other words, the bivariate p -box can be obtained by applying the imprecise version of Sklar's theorem (Proposition 4.111) with the product copula.

In particular, this also holds for other (stronger) conditions than factorisation also discussed in [52], such as the Kuznetsov property.

Note also that in our definition of the marginal coherent lower previsions $\underline{P}_X, \underline{P}_Y$ we have considered the natural extensions of their restrictions to cumulative sets; however,

the result still holds if we consider any other coherent extension, since in our use of the factorisation condition only the values in $A_{(x,y^*)}$, $A_{(x^*,y)}$ matter. We conclude then that, even if the independent natural extension and the strong product do not coincide in general [205, Section 9.3.4], they agree with respect to their associated bivariate p-box.

Interestingly, not all independent products induce the same p -box determined by the copula of the product:

Example 4.121 Consider $\Omega_1 = \Omega_2 = \{0, 1\}$ and let $\underline{F}_X = \underline{F}_Y$ be the marginal distribution functions given by $\underline{F}_X(0) = \underline{F}_Y(0) = 0.5$, $\underline{F}_X(1) = \underline{F}_Y(1) = 1$. They induce the marginal coherent lower previsions $\underline{P}_X, \underline{P}_Y$ given by

$$\underline{P}_X(f) = \min\{f(0), 0.5f(0) + 0.5f(1)\} \text{ and } \underline{P}_Y(g) = \min\{g(0), 0.5g(0) + 0.5g(1)\}$$

for every $f \in \mathcal{L}(\Omega_1)$, $g \in \mathcal{L}(\Omega_2)$. Their strong product is given by:

$$\underline{P}_X \boxtimes \underline{P}_Y := \min\{(0.25, 0.25, 0.25, 0.25), (0.5, 0, 0.5, 0), (0.5, 0.5, 0, 0), (1, 0, 0, 0)\}, \quad (4.20)$$

where in the above equation a vector (a, b, c, d) is used to denote the vector of probabilities $\{(P(0, 0), P(0, 1), P(1, 0), P(1, 1))\}$. Let \underline{P} be the coherent lower prevision given by

$$\underline{P} := \min\{(0.375, 0.125, 0.375, 0.125), (0.375, 0.375, 0.125, 0.125), (1, 0, 0, 0)\}.$$

Then the marginals of \underline{P} are also $\underline{P}_X, \underline{P}_Y$. Moreover, we see from Equation (4.20) that \underline{P} dominates $\underline{P}_X \boxtimes \underline{P}_Y$, and this allows us to deduce that \underline{P} is weakly coherent with both $\underline{P}_X(\cdot|\Omega_2), \underline{P}_Y(\cdot|\Omega_1)$: given a gamble f on $\Omega_1 \times \Omega_2$,

$$\underline{P}(G(f|\Omega_2)) \geq (\underline{P}_X \boxtimes \underline{P}_Y)(G(f|\Omega_2)) \geq 0,$$

whence in particular $\underline{P}(G(f|y)) = \underline{P}(G(fI_y|\Omega_2)) \geq 0$ for every $y \in \Omega_2$. And since \underline{P}_Y is the marginal of \underline{P} , it follows that we must have $\underline{P}(G(f|y)) = 0$: if it were $\underline{P}(G(f|y)) > 0$ then we would define the gamble g by $g(x', y') = f(x', y)$ and

$$0 = \underline{P}(g - \underline{P}_X(g)) \geq \sum_{y' \in \Omega_2} \underline{P}(G(g|y')) > 0,$$

a contradiction. Similarly, $\underline{P}(G(f|\Omega_1)) \geq 0$ and $\underline{P}(G(f|x)) = 0$ for every $x \in \Omega_1$. Applying [137, Theorem 1], we conclude that $\underline{P}, \underline{P}_X(\cdot|\Omega_1), \underline{P}_Y(\cdot|\Omega_1)$ are weakly coherent, and since $\underline{P}_X(\cdot|\Omega_2), \underline{P}_Y(\cdot|\Omega_1)$ are coherent because they are jointly coherent with $\underline{P}_X \boxtimes \underline{P}_Y$, we deduce from the reduction theorem [205, Theorem 7.1.5] that $\underline{P}, \underline{P}_X(\cdot|\Omega_2), \underline{P}_Y(\cdot|\Omega_1)$ are coherent. Thus, \underline{P} is an independent product. Its associated distribution function is given by

$$\underline{F}(0, 0) = 0.375, \quad \underline{F}(0, 1) = 0.5, \quad \underline{F}(1, 0) = 0.5, \quad \underline{F}(1, 1) = 1.$$

This differs from the bivariate distribution function \underline{F}' induced by $\underline{P}_X \boxtimes \underline{P}_Y$, which is the product of its marginals and which satisfies therefore $\underline{F}'(0, 0) = 0.25$. ♦

4.3.3 The role of imprecise copulas in the imprecise orders

Next we study how imprecise copulas can be used to express the relationship between imprecise stochastic dominance and statistical preference, that arise by using \succeq_{FSD} and \succeq_{SP} as the binary relation \succeq in Section 4.1. Afterwards, we shall study the role of imprecise copulas with respect to imprecise bivariate stochastic orders.

Univariate orders

We have seen in Section 3.2 that, although first degree stochastic dominance does not imply statistical preference in general (see Example 3.43), there are situations in which the implication holds (see Theorem 3.64), in terms of the marginal distributions of the variables and the copula that determines their joint distribution.

Given two random variables X and Y , let us denote by $\mathcal{C}_{X,Y}$ the set of copulas that make stochastic dominance imply statistical preference. Since the latter depends on the joint distribution of the random variables, it may be that X is preferred to Y when their joint distribution is determined by a copula C_1 and Y is preferred to X when it is determined by different copula C_2 .

In the imprecise framework, it is possible to establish the following connection between the imprecise stochastic dominance and statistical preference. We shall assume that we have imprecise information about the marginal distributions (that we model by means of p-boxes) and by the copula that links the marginal distributions into a joint (that we model by means of a set of copulas), in a manner similar to Proposition 4.111:

Proposition 4.122 *Consider a coherent lower prevision \underline{P} defined on the space product $\mathcal{X} \times \mathcal{Y}$ of two finite spaces that is factorising. Denote by $(\underline{E}, \overline{F})$ its associated bivariate p-box, that from Proposition 4.120 is built from the marginal p-boxes $(\underline{E}_X, \overline{F}_X)$ and $(\underline{E}_Y, \overline{F}_Y)$ using the product copula. Then, it holds that:*

$$(\underline{E}_X, \overline{F}_X) \succeq_{\text{FSD}_i} (\underline{E}_Y, \overline{F}_Y) \Rightarrow \mathcal{X} \succeq_{\text{SP}_i} \mathcal{Y}$$

for any $i = 1, \dots, 6$, where \mathcal{X} (respectively \mathcal{Y}) denotes the set of random variables whose cumulative distribution function belongs to $(\underline{E}_X, \overline{F}_X)$ ($(\underline{E}_Y, \overline{F}_Y)$, respectively).

Proof: We know from Proposition 4.120 that $(\underline{E}, \overline{F})$ is built by applying the product copula to their marginal p-boxes.

- $i = 1$: We know that for any $F_X \in (\underline{E}_X, \overline{F}_X)$ and $F_Y \in (\underline{E}_Y, \overline{F}_Y)$, $F_X \succeq_{\text{FSD}} F_Y$. Since they are coupled by the product copula, Theorem 3.44 implies $P_{F_X} \succeq_{\text{SP}} P_{F_Y}$. Thus, $X \succeq_{\text{SP}_1} Y$.

- $i = 2$: We know that there is $F_X^* \in (\underline{F}_X, \overline{F}_X)$ such that $F_X^* \succeq_{\text{FSD}} F_Y$ for any $F_Y \in (\underline{F}_Y, \overline{F}_Y)$. Since they are coupled by the product copula, Theorem 3.44 implies $P_{F_X^*} \succeq_{\text{SP}} P_{F_Y}$ for any $F_Y \in (\underline{F}_Y, \overline{F}_Y)$. Then, $X \succeq_{\text{SP}_2} Y$.
- $i = 3$: We know that for any $F_Y \in (\underline{F}_Y, \overline{F}_Y)$ there is $F_X \in (\underline{F}_X, \overline{F}_X)$ such that $F_X \succeq_{\text{FSD}} F_Y$. Then, for any P_{F_Y} , there is a P_{F_X} such that $F_X \succeq_{\text{FSD}} F_Y$, and consequently, the product copula links them, and by Theorem 3.44, $P_{F_X} \succeq_{\text{SP}} P_{F_Y}$.
- $i = 4$: We know that there are $F_X \in (\underline{F}_X, \overline{F}_X)$ and $F_Y \in (\underline{F}_Y, \overline{F}_Y)$ such that $F_X \succeq_{\text{FSD}} F_Y$. Then, consider P_{F_X} and P_{F_Y} . Since they are coupled by the product copula, Theorem 3.44 implies $P_{F_X} \succeq_{\text{SP}} P_{F_Y}$.
- The proof of cases $i = 5$ and $i = 6$ are similar to the one of cases $i = 2$ and $i = 3$. ■

Remark 4.123 *Although we may think that the previous result also holds when we build the joint bivariate p -box from the marginal p -boxes by means of a set of copulas $\mathcal{C} \subseteq \mathcal{C}_{X,Y}$, in the manner of Proposition 4.111, such a result does not seem to hold in general. The reason is that, as soon as one of the marginal p -boxes is imprecise (i.e., if its lower and the upper bounds do not coincide), we can find a distribution function inside the p -box associated with a neither continuous nor discrete random variable, and then, taking into account Theorem 3.64, we cannot assure the implication $\succeq_{\text{FSD}} \Rightarrow \succeq_{\text{SP}}$ unless we assume independence between the two p -boxes. ♦*

Bivariate orders

As we saw in Equation (2.6), univariate stochastic dominance can be expressed in terms of the comparison of expectations. It is also well-known that stochastic dominance can be expressed by means of the comparison of the survival distribution functions: given two random variables X and Y , their distribution functions are given by F_X and F_Y , and let $\tilde{F}_X(t) = P(X > t)$ and $\tilde{F}_Y = P(Y > t)$ denote their associated survival distribution functions. Then, it holds that:

$$F_X(t) = P(X \leq t) \leq P(Y \leq t) = F_Y(t) \Leftrightarrow \tilde{F}_X(t) = 1 - F_X(t) \geq 1 - F_Y(t) = \tilde{F}_Y. \quad (4.21)$$

Indeed, according to Equation (2.5), we have the following characterisations for first degree stochastic dominance:

$$\begin{aligned} X \succeq_{\text{FSD}} Y &\Leftrightarrow F_X(t) \leq F_Y(t) \text{ for any } t \\ &\Leftrightarrow E[u(X)] \geq E[u(Y)] \text{ for any increasing } u \\ &\Leftrightarrow \tilde{F}_X(t) \geq \tilde{F}_Y(t) \text{ for any } t. \end{aligned}$$

In the bivariate case, the survival distribution functions are not related to the distribution functions as in Equation (4.21), since $P(X > t_1, Y > t_2) \neq 1 - P(X \leq t_1, Y \leq t_2)$. Then,

these three conditions are not equivalent, and they generate three different stochastic orders:

Definition 4.124 Let (X_1, X_2) and (Y_1, Y_2) be two random vectors with bivariate distribution functions F_{X_1, X_2} and F_{Y_1, Y_2} . We say that:

- (X_1, X_2) stochastically dominates (Y_1, Y_2) , and denote it $(X_1, X_2) \succeq_{\text{FSD}} (Y_1, Y_2)$, if $E[u(X_1, X_2)] \geq E[u(Y_1, Y_2)]$ for any increasing $u : \mathbb{R}^2 \rightarrow \mathbb{R}$.
- (X_1, X_2) is preferred to (Y_1, Y_2) with respect to the upper orthant order, and denote it $(X_1, X_2) \succeq_{\text{uo}} (Y_1, Y_2)$, if $\tilde{F}_{X_1, X_2}(t) \geq \tilde{F}_{Y_1, Y_2}(t)$ for any $t \in \mathbb{R}^2$.
- (X_1, X_2) is preferred to (Y_1, Y_2) with respect to the lower orthant order, and denote it $(X_1, X_2) \succeq_{\text{lo}} (Y_1, Y_2)$, if $F_{X_1, X_2}(t) \leq F_{Y_1, Y_2}(t)$ for any $t \in \mathbb{R}^2$.

These three orders are equivalent in the univariate case, but not in the bivariate. Next theorem describe the relationships between these three orders:

Theorem 4.125 ([139, Theorem 3.3.2]) If $X \succeq_{\text{FSD}} Y$, then $X \succeq_{\text{lo}} Y$ and $X \succeq_{\text{uo}} Y$. In addition, there is no implication between the lower and the upper orthant orders.

In Remark 4.127 we will give an example where the lower and the upper orthant orders are not equivalent.

Since any copula C is in particular a bivariate distribution function on $[0, 1] \times [0, 1]$, the previous orders can also be applied to the comparison of copulas. Taking this into account, we can establish the following result, that links the comparison of bivariate p-boxes with the comparison of their associated marginal p-boxes.

Proposition 4.126 Let $(\underline{F}_{X_1}, \overline{F}_{X_1}), (\underline{F}_{X_2}, \overline{F}_{X_2}), (\underline{F}_{Y_1}, \overline{F}_{Y_1})$ be univariate p-boxes and $(\underline{F}_{Y_2}, \overline{F}_{Y_2})$ and the set of copulas \mathcal{C}_X and \mathcal{C}_Y . Let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ be the bivariate p-boxes given by:

$$\begin{aligned} (\underline{F}_X, \overline{F}_X) &:= \{C(F_{X_1}, F_{X_2}) : C \in \mathcal{C}_X, F_{X_1} \in (\underline{F}_{X_1}, \overline{F}_{X_1}), F_{X_2} \in (\underline{F}_{X_2}, \overline{F}_{X_2})\} \\ (\underline{F}_Y, \overline{F}_Y) &:= \{C(F_{Y_1}, F_{Y_2}) : C \in \mathcal{C}_Y, F_{Y_1} \in (\underline{F}_{Y_1}, \overline{F}_{Y_1}), F_{Y_2} \in (\underline{F}_{Y_2}, \overline{F}_{Y_2})\}. \end{aligned}$$

Then, it holds that:

$$\left. \begin{aligned} (\underline{F}_{X_1}, \overline{F}_{X_1}) \succeq_{\text{FSD}_i} (\underline{F}_{Y_1}, \overline{F}_{Y_1}) \\ (\underline{F}_{X_2}, \overline{F}_{X_2}) \succeq_{\text{FSD}_i} (\underline{F}_{Y_2}, \overline{F}_{Y_2}) \\ \mathcal{C}_X \succeq_{\text{lo}_i} \mathcal{C}_Y \end{aligned} \right\} \Rightarrow (\underline{F}_X, \overline{F}_X) \succeq_{\text{lo}_i} (\underline{F}_Y, \overline{F}_Y)$$

for $i = 1, \dots, 6$.

Proof:

($i = 1$) We know that:

$$\begin{aligned} \forall F_{X_1} \in (\underline{F}_{X_1}, \overline{F}_{X_1}), F_{Y_1} \in (\underline{F}_{Y_1}, \overline{F}_{Y_1}), F_{X_1} \leq F_{Y_1}. \\ \forall F_{X_2} \in (\underline{F}_{X_2}, \overline{F}_{X_2}), F_{Y_2} \in (\underline{F}_{Y_2}, \overline{F}_{Y_2}), F_{X_2} \leq F_{Y_2}. \\ \forall C_X \in \mathcal{C}_X, C_Y \in \mathcal{C}_Y, C_X \leq C_Y. \end{aligned}$$

Consider $F_X \in (\underline{F}_X, \overline{F}_X)$ and $F_Y \in (\underline{F}_Y, \overline{F}_Y)$. They can be expressed in the following way: $F_X(x, y) = C_X(F_{X_1}(x), F_{X_2}(y))$ and $F_Y(x, y) = C_Y(F_{Y_1}(x), F_{Y_2}(y))$. Then:

$$\begin{aligned} F_X(x, y) = C_X(F_{X_1}(x), F_{X_2}(y)) \leq C_X(F_{Y_1}(x), F_{Y_2}(y)) \\ \leq C_Y(F_{Y_1}(x), F_{Y_2}(y)) = F_Y(x, y). \end{aligned}$$

($i = 2$) We know that:

$$\begin{aligned} \exists F_{X_1}^* \in (\underline{F}_{X_1}, \overline{F}_{X_1}) \text{ such that } F_{X_1}^* \leq F_{Y_1} \quad \forall F_{Y_1} \in (\underline{F}_{Y_1}, \overline{F}_{Y_1}). \\ \exists F_{X_2}^* \in (\underline{F}_{X_2}, \overline{F}_{X_2}) \text{ such that } F_{X_2}^* \leq F_{Y_2} \quad \forall F_{Y_2} \in (\underline{F}_{Y_2}, \overline{F}_{Y_2}). \\ \exists C_X^* \in \mathcal{C}_X \text{ such that } C_X^* \leq C_Y \quad \forall C_Y \in \mathcal{C}_Y. \end{aligned}$$

Consider $F_X(x, y) := C_X^*(F_{X_1}^*(x), F_{X_2}^*(y))$, and let us see that $F_X \leq F_Y$ for any $F_Y(x, y) = C_Y(F_{Y_1}(x), F_{Y_2}(y))$:

$$\begin{aligned} F_X(x, y) = C_X^*(F_{X_1}^*(x), F_{X_2}^*(y)) \leq C_X^*(F_{Y_1}(x), F_{X_2}(y)) \\ \leq C_Y(F_{Y_1}(x), F_{X_2}(y)) = F_Y(x, y). \end{aligned}$$

($i = 3$) We know that:

$$\begin{aligned} \forall F_{Y_1} \in (\underline{F}_{Y_1}, \overline{F}_{Y_1}), \exists F_{X_1} \in (\underline{F}_{X_1}, \overline{F}_{X_1}) \text{ such that } F_{X_1} \leq F_{Y_1}. \\ \forall F_{Y_2} \in (\underline{F}_{Y_2}, \overline{F}_{Y_2}), \exists F_{X_2} \in (\underline{F}_{X_2}, \overline{F}_{X_2}) \text{ such that } F_{X_2} \leq F_{Y_2}. \\ \forall C_Y \in \mathcal{C}_Y \exists C_X \in \mathcal{C}_X \text{ such that } C_X \leq C_Y. \end{aligned}$$

Consider $F_Y(x, y) = C_Y(F_{Y_1}(x), F_{Y_2}(y))$, and let us check that there is F_X such that $F_X \leq F_Y$. We define $F_X(x, y) = C_X(F_{X_1}(x), F_{X_2}(y))$ such that $C_X \leq C_Y$, $F_{X_1} \leq F_{Y_1}$ and $F_{X_2} \leq F_{Y_2}$. Then:

$$\begin{aligned} F_X(x, y) = C_X(F_{X_1}(x), F_{X_2}(y)) \leq C_X(F_{Y_1}(x), F_{Y_2}(y)) \\ \leq C_Y(F_{Y_1}(x), F_{Y_2}(y)) = F_Y(x, y). \end{aligned}$$

($i = 4$) We know that:

$$\begin{aligned} \exists F_{X_1}^* \in (\underline{F}_{X_1}, \overline{F}_{X_1}), F_{Y_1}^* \in (\underline{F}_{Y_1}, \overline{F}_{Y_1}) \text{ such that } F_{X_1}^* \leq F_{Y_1}^*. \\ \exists F_{X_2}^* \in (\underline{F}_{X_2}, \overline{F}_{X_2}), F_{Y_2}^* \in (\underline{F}_{Y_2}, \overline{F}_{Y_2}) \text{ such that } F_{X_2}^* \leq F_{Y_2}^*. \\ \exists C_X^* \in \mathcal{C}_X, C_Y^* \in \mathcal{C}_Y \text{ such that } C_X^* \leq C_Y^*. \end{aligned}$$

Consider $F_X(x, y) = C_X^*(F_{X_1}^*(x), F_{X_2}^*(y))$ and $F_Y(x, y) = C_Y^*(F_{Y_1}^*(x), F_{Y_2}^*(y))$. It holds that $F_X \leq F_Y$:

$$F_X(x, y) = C_X^*(F_{X_1}^*(x), F_{X_2}^*(y)) \leq C_X^*(F_{Y_1}^*(x), F_{Y_2}^*(y)) \leq C_Y^*(F_{Y_1}^*(x), F_{Y_2}^*(y)) = F_Y(x, y).$$

($i = 5, i = 6$) The proof of these two cases is analogous to that of $i = 2$ and $i = 3$, respectively. ■

Remark 4.127 Note that under the hypotheses of Proposition 4.126 we do not necessarily have that $(\underline{F}_X, \overline{F}_X) \succeq_{\text{uo}_i} (\underline{F}_Y, \overline{F}_Y)$. To see this, consider the following probability mass functions (see [139, Example 3.3.3]):

$X_2 \backslash X_1$	0	1	2	$Y_2 \backslash Y_1$	0	1	2
0	0	0	$\frac{1}{8}$	0	$\frac{1}{4}$	$\frac{1}{4}$	0
1	$\frac{1}{4}$	$\frac{1}{4}$	0	1	0	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{4}$	$\frac{1}{8}$	0	2	$\frac{1}{4}$	0	0

Then, $(X_1, X_2) \succeq_{\text{lo}} (Y_1, Y_2)$ since $F_{X_1, X_2} \leq F_{Y_1, Y_2}$. However, $(X_1, X_2) \not\succeq_{\text{uo}} (Y_1, Y_2)$, since:

$$\tilde{F}_X(1, 0) = P(X_1 > 1, X_2 > 0) = 0 < \frac{1}{8} = P(Y_1 > 1, Y_2 > 0) = \tilde{F}_Y(1, 0).$$

This example also shows that under the assumptions of Proposition 4.126 it does not necessarily hold that $(X_1, X_2) \succeq_{\text{FSD}} (Y_1, Y_2)$; otherwise, we would deduce from Theorem 4.125 that $(X_1, X_2) \succeq_{\text{uo}} (Y_1, Y_2)$, a contradiction with the example above. ◆

A result similar to Proposition 4.126 can be established when we consider the upper instead of the lower orthant order:

Proposition 4.128 Let $(\underline{F}_{X_1}, \overline{F}_{X_1}), (\underline{F}_{X_2}, \overline{F}_{X_2}), (\underline{F}_{Y_1}, \overline{F}_{Y_1})$ be univariate p -boxes and $(\underline{F}_{Y_2}, \overline{F}_{Y_2})$ and the set of copulas \mathcal{C}_X and \mathcal{C}_Y . Let $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ be the bivariate p -boxes given by:

$$\begin{aligned} (\underline{F}_X, \overline{F}_X) &:= \{C(F_{X_1}, F_{X_2}) : C \in \mathcal{C}_X, F_{X_1} \in (\underline{F}_{X_1}, \overline{F}_{X_1}), F_{X_2} \in (\underline{F}_{X_2}, \overline{F}_{X_2})\} \\ (\underline{F}_Y, \overline{F}_Y) &:= \{C(F_{Y_1}, F_{Y_2}) : C \in \mathcal{C}_Y, F_{Y_1} \in (\underline{F}_{Y_1}, \overline{F}_{Y_1}), F_{Y_2} \in (\underline{F}_{Y_2}, \overline{F}_{Y_2})\}. \end{aligned}$$

Then, it holds that:

$$\left. \begin{aligned} (\underline{F}_{X_1}, \overline{F}_{X_1}) &\succeq_{\text{FSD}_i} (\underline{F}_{Y_1}, \overline{F}_{Y_1}) \\ (\underline{F}_{X_2}, \overline{F}_{X_2}) &\succeq_{\text{FSD}_i} (\underline{F}_{Y_2}, \overline{F}_{Y_2}) \\ \mathcal{C}_X &\succeq_{\text{uo}_i} \mathcal{C}_Y \end{aligned} \right\} \Rightarrow (\underline{F}_X, \overline{F}_X) \succeq_{\text{uo}_i} (\underline{F}_Y, \overline{F}_Y),$$

for $i = 1, \dots, 6$.

The proof of this result is analogous to the one of Proposition 4.126, and therefore omitted.

Natural extension and independent product

To conclude this section, we consider the particular cases where the bivariate p-boxes are made by means of the natural extension or a factorising product.

By Proposition 4.117, the natural extension of two marginal p-boxes $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ is given by:

$$\underline{F}(x, y) = C_L(\underline{F}_X(x), \underline{F}_Y(y)) \text{ and } \overline{F}(x, y) = C_M(\overline{F}_X(x), \overline{F}_Y(y)). \quad (4.22)$$

This allows us to prove the following result:

Corollary 4.129 *Consider marginal p-boxes $(\underline{F}_{X_1}, \overline{F}_{X_1})$, $(\underline{F}_{X_2}, \overline{F}_{X_2})$ and $(\underline{F}_{Y_1}, \overline{F}_{Y_1})$ and $(\underline{F}_{Y_2}, \overline{F}_{Y_2})$. Let $(\underline{F}_X, \overline{F}_X)$ (respectively, $(\underline{F}_Y, \overline{F}_Y)$) denote the natural extension of the p-boxes $(\underline{F}_{X_1}, \overline{F}_{X_1})$ and $(\underline{F}_{X_2}, \overline{F}_{X_2})$ (respectively, $(\underline{F}_{Y_1}, \overline{F}_{Y_1})$, $(\underline{F}_{Y_2}, \overline{F}_{Y_2})$) by means of Equation (4.22). Then:*

$$\left. \begin{array}{l} (\underline{F}_{X_1}, \overline{F}_{X_1}) \succeq_{\text{FSD}_i} (\underline{F}_{Y_1}, \overline{F}_{Y_1}) \\ (\underline{F}_{X_2}, \overline{F}_{X_2}) \succeq_{\text{FSD}_i} (\underline{F}_{Y_2}, \overline{F}_{Y_2}) \end{array} \right\} \Rightarrow (\underline{F}_X, \overline{F}_X) \succeq_{\text{lo}_i} (\underline{F}_Y, \overline{F}_Y)$$

for $i = 2, \dots, 6$.

Proof: The result follows immediately from Proposition 4.126. ■

To see that the result does not hold for \succeq_{lo_1} , consider the following example.

Example 4.130 *For $j = 1, 2$, let $\underline{F}_{X_j} = \overline{F}_{X_j} = \underline{F}_{Y_j} = \overline{F}_{Y_j}$ be the distribution function associated with a uniform distribution on $[0, 1]$, and let us denote it by F . Then, trivially:*

$$(\underline{F}_{X_j}, \overline{F}_{X_j}) \succeq_{\text{FSD}_1} (\underline{F}_{Y_j}, \overline{F}_{Y_j}) \text{ for } j = 1, 2.$$

To see that $(\underline{F}_X, \overline{F}_X) \not\succeq_{\text{lo}_1} (\underline{F}_Y, \overline{F}_Y)$, it suffices to note that $C_M(F, F) \in (\underline{F}_X, \overline{F}_X)$ and $C_L(F, F) \in (\underline{F}_Y, \overline{F}_Y)$, and:

$$C_M(F(0.5), F(0.5)) = C_M(0.5, 0.5) = 0.5 > 0 = C_L(0.5, 0.5) = C_L(F(0.5), F(0.5)). \blacklozenge$$

We also saw in Proposition 4.120 that the bivariate p-box associated with a factorising coherent lower probability is obtained applying the product copula to the two marginal p-boxes. This fact allows us to simplify Propositions 4.126 and 4.128:

Corollary 4.131 *Consider two factorising coherent lower probabilities \underline{P}_X and \underline{P}_Y defined on $\mathcal{X} \times \mathcal{Y}$, where both sets are finite. Denote by $(\underline{F}_X, \overline{F}_X)$ and $(\underline{F}_Y, \overline{F}_Y)$ their associated bivariate p-boxes, that from Proposition 4.120 can be obtained by applying the product copula to their respective marginal distributions represented by the p-boxes*

$(\underline{F}_{X_1}, \overline{F}_{X_1})$, $(\underline{F}_{X_2}, \overline{F}_{X_2})$ and $(\underline{F}_{Y_1}, \overline{F}_{Y_1})$ and $(\underline{F}_{Y_2}, \overline{F}_{Y_2})$, respectively. Then, it holds that:

$$\left. \begin{array}{l} (\underline{F}_{X_1}, \overline{F}_{X_1}) \succeq_{\text{FSD}_i} (\underline{F}_{Y_1}, \overline{F}_{Y_1}) \\ (\underline{F}_{X_2}, \overline{F}_{X_2}) \succeq_{\text{FSD}_i} (\underline{F}_{Y_2}, \overline{F}_{Y_2}) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (\underline{F}_X, \overline{F}_X) \succeq_{\text{lo}_i} (\underline{F}_Y, \overline{F}_Y) \\ (\underline{F}_X, \overline{F}_X) \succeq_{\text{uo}_i} (\underline{F}_Y, \overline{F}_Y) \end{array} \right.$$

Proof: We have seen in Proposition 4.120 that the bivariate p-box associated with a factorising coherent lower probability is made by considering the product copula applied to the marginal p-boxes. Then, this result is a particular case of Propositions 4.126 and 4.128. ■

4.4 Applications

To conclude the chapter, we give some possible applications of the extension of stochastic orders to an imprecise framework. We start with two possible applications of imprecise stochastic dominance: the comparison of Lorenz Curves and that of cancer survival rates. Lorenz Curves are a well-known economic tool that measure how the wealth of a population is distributed. Since Lorenz Curves can be seen as distribution functions, we can compare them by means of stochastic dominance. Furthermore, in some cases the economical analysis is made for geographical regions that comprise several countries, like for example Nordic countries, Southern Europe, American, ... Then, we can use the imprecise stochastic dominance to compare the sets of Lorenz Curves associated with these groups of countries. On the other hand, some kind of cancer sites can also be grouped into Digestive, Respiratory, Reproductive or Other. Then, it is possible to compare the survival rates of the group of cancer by comparing their associated set of mortality rates, that can be expressed as distribution functions. Then, also the imprecise stochastic dominance could be applied.

Afterwards, we focus on a Multi-Criteria Decision Making problem, where it is possible to find imprecision in the utilities or in the beliefs. This allows us to illustrate how both the imprecise stochastic dominance and statistical preference can be used, as well as the strong and weak dominance introduced in Section 4.2.2.

4.4.1 Comparison of Lorenz curves

As we mentioned in Section 2.1.1, the notion of stochastic dominance has been applied in many different contexts. One of the most interesting is in the field of social welfare [3, 117, 190], for comparing *Lorenz curves*. They are a graphical representation of the cumulative distribution function of the wealth: the elements of the population are ordered according to it, and the curve shows, for the bottom x% elements, what percentage y% of

Country-year	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1
Australia-1994	5.9	12.01	17.2	23.57	41.32
Canada-2000	7.2	12.73	17.18	22.95	39.94
China-2005	5.73	9.8	14.66	22	47.81
Finland-2000	9.62	14.07	17.47	22.14	36.7
FYR Macedonia-2000	9.02	13.45	17.49	22.61	37.43
Greece-2000	6.74	11.89	16.84	23.04	41.49
India-2005	8.08	11.27	14.94	20.37	45.34
Japan-1993	10.58	14.21	17.58	21.98	35.65
Maldives-2004	6.51	10.88	15.71	22.66	44.24
Norway-2000	9.59	13.96	17.24	21.98	37.23
Sweden-2000	9.12	13.98	17.57	22.7	36.63
USA-2000	5.44	10.68	15.66	22.4	45.82

Table 4.2: Quintiles of the Lorenz Curves associated with different countries.

the total wealth they have. Hence, the Lorenz curve can be used as a measure of equality: the closest the curve is to the straight line, the more equal the associated society is.

If we have the Lorenz curves of two different countries, we can compare them by means of stochastic dominance: if one of them is dominated by the other, the closest to the straight line will be associated with a more equal society, and will therefore be considered preferable. In this section, we are going to use our extensions of stochastic dominance to compare sets of Lorenz curves associated with countries in different areas of the world. We shall consider the Lorenz curves associated with the quintiles of the empirical distribution functions. Table 4.2 provides the wealth in each of the quintiles (Source data: World Bank database. <http://timetric.com/dataset/worldbank>):

To make the comparison by means of the extensions of stochastic dominance clearer, we are going to consider the cumulative distribution from the richest to the poorest group: in this way, we will always obtain a curve which is above the straight line, and it will comply with our idea of considering preferable the smallest distribution function. If we apply this to the data in Table 4.2, we obtain the data of Table 4.3.

We are going to group these countries by continents/regions:

- Group 1: China, Japan, India.
- Group 2: Finland, Norway, Sweden.
- Group 3: Canada, USA.
- Group 4: FYR Macedonia, Greece.

Country-year	F(0.2)	F(0.4)	F(0.6)	F(0.8)	F(1)
Australia-1994	41.32	64.89	82.09	94.1	100
Canada-2000	39.94	62.89	80.07	92.8	100
China-2005	47.81	69.81	84.47	94.27	100
Finland-2000	36.7	58.84	76.31	90.38	100
FYR Macedonia-2000	37.43	60.04	77.53	90.98	100
Greece-2000	41.49	64.53	81.37	93.26	100
India-2005	45.34	65.71	80.65	91.92	100
Japan-1993	35.65	57.63	75.21	89.42	100
Maldives-2004	44.24	66.9	82.61	93.49	100
Norway-2000	37.23	59.21	76.45	90.41	100
Sweden-2000	36.63	59.33	76.9	90.88	100
USA-2000	45.82	68.22	83.88	94.56	100

Table 4.3: Cumulative distribution functions associated with the Lorenz Curves of the countries.

	Group 1	Group 2	Group 3	Group 4	Group 5
Group 1	$\equiv_{FSD_{2,5}}$	\succ_{FSD_2}	\succ_{FSD_2}	\succ_{FSD_2}	\succ_{FSD_2}
Group 2	\succ_{FSD_5}	$\equiv_{FSD_{3,6}}$	\succ_{FSD_1}	\succ_{FSD_1}	\succ_{FSD_1}
Group 3	\equiv_{FSD_4}		$\equiv_{FSD_{2,5}}$	\succ_{FSD_2}	\succ_{FSD_2}
Group 4	\succ_{FSD_5}		\succ_{FSD_5}	$\equiv_{FSD_{3,6}}$	$\succ_{FSD_{3,6}}$
Group 5	\succ_{FSD_5}		\succ_{FSD_5}		$\equiv_{FSD_{3,6}}$

Table 4.4: Result of the comparison of the regions by means of the imprecise stochastic dominance.

- Group 5: Australia, Maldives.

The relationships between these groups are summarised in Table 4.4.

This means for instance that the set of distribution functions in the first group strictly dominates the second group according to definition (FSD_2), while the second group strictly dominates the first group according to definition (FSD_5). This is because the best country in the first group (Japan) stochastically dominates all the countries in the second group, but the worst (China) is stochastically dominated by all countries in the second group. This, together with Proposition 4.3, implies that the first group strictly dominates the second according to (FSD_3), is strictly dominated by the second according to (FSD_6), that they are indifferent according to (FSD_4) and incomparable according to (FSD_1).

Similar considerations hold for the other pairwise comparisons. For instance, group

4 strictly dominates group 5 according to $(FSD_3), (FSD_6)$, but it does not dominate it according to $(FSD_2), (FSD_5)$. This also shows that conditions (FSD_2) and (FSD_3) are not equivalent (and similarly for (FSD_5) and (FSD_6)).

The cells where we have left a blank space mean that no dominance relationship is satisfied: for instance, group 3 does not dominate group 2 according to any of the definitions.

Since all the groups have more than one element, they will not satisfy (FSD_1) when comparing them to themselves. It follows from Remark 4.31 that they are always indifferent to themselves according to $(FSD_3), (FSD_4)$ and (FSD_6) ; they are indifferent to themselves according to (FSD_2) when they have a best-case-scenario (as it is the case for groups 1 and 3), and indifferent according to (FSD_5) when they have a worst-case scenario (as it is the case again for groups 1 and 3), and incomparable according to these definitions in the other cases.

Note that we can also use the above data to illustrate some of the results in this paper: for instance, we saw in Remark 4.9 that condition (FSD_2) is transitive, and in the table above we see that group 1 is preferred to group 3 according to (FSD_2) and group 3 is preferred to group 4 according to (FSD_2) : this allows us to infer immediately that group 1 is preferred to group 4 according to this condition. The comparison of the first two groups is an instance of Proposition 4.32, because the p-box induced by the first group is strictly more imprecise (i.e., it has a smaller lower cumulative distribution and a greater upper cumulative distribution function) than that of the second group.

Remark 4.132 *In economy, the Gini Index is a well-known inequality measure that express how the incomes of a population are shared. It takes values between 0 and 1, where a Gini Index of 0 means perfect equality for the incomes of the people, while a Gini Index of 1 express a total inequality in the incomes. Thus, the greater the Gini Index is, the more inequality the incomes of a population are.*

The Gini Index is quite related to Lorenz curves: given a Lorenz Curve F , that express the distribution function of the wealth of a population (a country, a region,...), its associated Gini index is defined by:

$$G = 2 \int_0^{100} (x - F(x))dx.$$

Thus, the closer the Lorenz curve is to the straight $y = x$, the smaller the Gini index is.

In the imprecise framework, if we are working with a p-box that represents the Lorenz curve, we can compute the lower and the upper Gini Indexes, that are a lower and an upper bound of the Gini Index, simply by considering the Gini indexes of the upper and the lower bounds of the p-box. Then, for any p-box $(\underline{F}, \overline{F})$ representing Lorenz curve F we obtain a Gini index given in an interval form: $[\underline{G}, \overline{G}]$, where \underline{G} is the Gini index associated with \overline{F} and \overline{G} is the Gini index associated with \underline{F} . Then, in order to compare

the Gini intervals associated with two imprecise Lorenz curves, it is possible to consider the usual orderings for real intervals (see for instance [69, 78]).◆

4.4.2 Comparison of cancer survival rates

According to [28], long-term cancer survival rates have substantially improved in the past decades. However, there are still some kinds of cancer whose survival rates could clearly be improved. Here, we use the survival rates of different cancer sites given in [28]. These can be grouped in Digestive, Respiratory, Reproductive and Other, and we shall compare the survival rates of these types applying imprecise stochastic dominance.

Table 4.5 shows the survival rates of different cancer sites (see [28]).

Note that it is possible to transform the survival rates of Table 4.5 into cumulative distribution functions. In this case, we assume the distribution functions to be defined in the interval $[0,100]$, and we impose the condition $F(100) = 1$, that means that the survival rate after 100 years of the cancer diagnostic is zero. The results are showed in Table 4.6.

These cancer sites can be grouped as follows:

Digestives Colon (C), Rectum (R), Oral cavity and pharynx (OCP), Stomach (S), Oesophagus (O), Liver and intrahepatic bile duct (LIBD), Pancreas (P).

Respiratory Larynx (L), Lung and bronchus (LB).

Reproductive Prostate (Pr), Testis (T), Breast (B), Cervix uteri (CU), Corpus uteri and uterus (CUU), Ovary (Ov).

Other Melanomas (M), Urinary bladder (UB), Kidney and renal pelvis (KRP), Brain and other nervous system (BNS), Thyroid (Th), Hodgkin's disease (HD), Non-Hodgkin lymphomas (NHL), Leukaemias (L).

Let us compare these kinds of cancer by means of the imprecise stochastic dominance. Note that in this case, given two distribution functions F_1 and F_2 that represent the mortality rates of two cancer sites, $F_1 \succeq_{\text{FSD}} F_2$ means that the cancer F_1 is less deadly than the cancer F_2 , or equivalently, that the cancer F_1 has a greater survival rate than the cancer F_2 .

First of all, note that Pancreas (P) is the worst cancer with respect to stochastic dominance, since $F < F_P$ for any other distribution function F . This implies that Digestive is \succ_{FSD_5} dominated by the other three groups, and then, from a pessimistic point of view, digestive cancers are the worst. Furthermore, Prostate and Thyroid cancers are less deadly than any of the digestive cancers, and then both Reproductive and Other groups

Cancer site	Relative survival rate, %			
	1 year	4 years	7 years	10 years
Colon	80.7	65.6	60.5	58.2
Rectum	86.3	68.2	61.2	57.9
Oral cavity and pharynx	82.9	63.0	56.1	50.2
Stomach	49.0	27.0	22.9	20.8
Oesophagus	43.4	17.9	13.8	11.8
Liver and intrahepatic bile duct	34.5	15.2	11.0	9.2
Pancreas	23.0	6.2	4.5	3.8
Larynx	85.9	66.3	57.0	49.6
Lung and bronchus	41.2	17.5	13.0	10.5
Prostate*	99.6	98.6	97.9	97.0
Testis*	97.8	95.7	95.4	95.0
Breast**	97.5	90.4	85.8	82.6
Cervix uteri**	88.0	72.3	68.3	66.1
Corpus uteri and uterus**	92.4	83.9	81.5	80.3
Ovary**	74.9	48.5	38.8	35.0
Melanomas	97.3	92.2	90.3	89.5
Urinary bladder	90.1	80.9	76.4	72.7
Kidney and renal pelvis	80.8	69.3	63.8	59.4
Brain and other nervous system	56.4	35.1	30.6	27.9
Thyroid	97.6	96.9	96.3	95.9
Hodgkin's disease	92.4	85.8	82.2	79.6
Non-Hodgkin lymphomas	77.2	65.1	59.0	54.3
Leukaemias	70.2	55.0	48.3	43.8

Table 4.5: Estimation of relative survival rates by cancer site. The rates are derived from SEER 1973-98 database, all ethnic groups, both sexes (except (*), only for male, and (**)) for female). [191].

\succ_{FSD_2} dominates Digestive. However, Digestive and Respiratory are incomparable with respect to (FSD_2) and (FSD_3) , and they are equivalent with respect to (FSD_4) , since $F_P > F_{LB} > F_C$. Also Digestive is (FSD_4) equivalent to Reproductive and Other groups, since $F_P > F_{Ov} > F_C$ and $F_P > F_{BNS} > F_C$.

Since Lung and Brounch cancer has a greater mortality than any Reproductive cancer, Respiratory is \succ_{FSD_5} dominated by Reproductive group. Furthermore, they are not comparable with respect to (FSD_2) and indifferent with respect to (FSD_4) since $F_L < F_{Ov} < F_{LL}$.

Finally, since Brain and other nervous system cancer is stochastically dominated by any Reproductive cancer, Reproductive \succ_{FSD_5} dominates Other group, and they are

	Cumulative distribution functions			
	$F(1)$	$F(4)$	$F(7)$	$F(10)$
Cancer site				
Colon	0.193	0.344	0.395	0.418
Rectum	0.137	0.318	0.388	0.421
Oral cavity and pharynx	0.171	0.370	0.439	0.498
Stomach	0.510	0.730	0.771	0.792
Oesophagus	0.566	0.821	0.862	0.882
Liver and intrahepatic bile duct	0.655	0.846	0.890	0.908
Pancreas	0.770	0.938	0.955	0.962
Larynx	0.141	0.337	0.430	0.504
Lung and bronchus	0.588	0.825	0.870	0.895
Prostate	0.004	0.014	0.021	0.030
Testis	0.022	0.043	0.046	0.050
Breast	0.025	0.096	0.142	0.174
Cervix uteri	0.120	0.277	0.317	0.339
Corpus uteri and uterus	0.076	0.161	0.185	0.197
Ovary	0.251	0.515	0.612	0.650
Melanomas	0.027	0.078	0.097	0.105
Urinary bladder	0.099	0.191	0.236	0.273
Kidney and renal pelvis	0.192	0.307	0.362	0.406
Brain and other nervous system	0.436	0.649	0.694	0.721
Thyroid	0.024	0.031	0.037	0.041
Hodgkin's disease	0.076	0.142	0.178	0.204
Non-Hodgkin lymphomas	0.228	0.349	0.410	0.457
Leukaemias	0.298	0.450	0.517	0.562

Table 4.6: Estimation of relative mortality rates by cancer site.

equivalent with respect to (FSD_4) since $F_M < F_{CU} < F_{BNS}$.

The results are depicted in Table 4.7.

Thus, according to our results, Digestive cancer seems to be the group with a greater mortality rate, while Reproductive cancer seems to be the least deadly.

4.4.3 Multiattribute decision making

In this section, we shall illustrate the extension of statistical preference to a context of imprecision by means of an application to decision making. We shall consider two different scenarios: on the one hand, we shall compare two alternatives in a context of

	Digestive	Respiratory	Reproductive	Other
Digestive	\equiv_{FSD_5}	\equiv_{FSD_4}	\equiv_{FSD_4}	\equiv_{FSD_4}
Respiratory	\succ_{FSD_5}	$\equiv_{\text{FSD}_{2,5}}$		\equiv_{FSD_4}
Reproductive	$\succ_{\text{FSD}_{2,5}}$	\succ_{FSD_5}	\equiv_{FSD_5}	\succ_{FSD_5}
Other	$\succ_{\text{FSD}_{2,5}}$	\succ_{FSD_2}	\equiv_{FSD_4}	$\equiv_{\text{FSD}_{2,5}}$

Table 4.7: Result of the comparison of the different groups of cancer by means of the imprecise stochastic dominance.

imprecise information about their utilities or probabilities, by means of the results in Sections 4.2.1 and 4.2.2; on the other hand, we shall consider the comparison of two sets of alternatives, by means of the techniques established in Section 4.1. Our running example throughout this section is based on [118, Section 4].

A decision problem with uncertain beliefs

Consider a decision problem where we must choose between n alternatives a_1, \dots, a_n , whose rewards depend on the values of the states of nature, $\theta_1, \dots, \theta_m$, which hold with certain probabilities $P(\theta_1), \dots, P(\theta_m)$.

Let us start by assuming that there is uncertainty about these probabilities, that we model by means of a set of probability measures \mathcal{P} . Then, we shall compare any two alternatives by means of the concepts of weak and strong \mathcal{P} -preference we have considered in Section 4.2.2.

Example 4.133 *A company must choose where to invest its money. The alternatives are: a_1 -a computer company; a_2 -a car company; a_3 -a fast food company. The rewards associated with the investment depend on an attribute c_1 : “economic evolution”, which may take the values θ_1 -“very good”, θ_2 -“good”, θ_3 -“normal” or θ_4 -“bad”. The probabilities of each of these states are expressed by means of an interval. The rewards associated with any combination (alternative, state) are expressed in a linguistic scale, with values $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ (very poor, poor, slightly poor, normal, slightly good, good, very good). The available information is summarised in the following table:*

	θ_1	θ_2	θ_3	θ_4
	[0.1, 0.4]	[0.2, 0.7]	[0.3, 0.4]	[0.1, 0.4]
a_1	s_4	s_3	s_3	s_2
a_2	s_5	s_4	s_4	s_2
a_3	s_2	s_3	s_5	s_4

Hence, the set \mathcal{P} of probability measures for our beliefs is given by

$$\mathcal{P} = \{(p_1, p_2, p_3, p_4) : p_1 + p_2 + p_3 + p_4 = 1, \\ p_1 \in [0.1, 0.4], p_2 \in [0.2, 0.7], p_3 \in [0.3, 0.4], p_4 \in [0.1, 0.4]\}.$$

Since the rewards are expressed in a qualitative scale, we are going to compare the different alternatives by means of statistical preference. We obtain that:

$$Q(a_1, a_2) = \frac{1}{2}p_4 \in [0.05, 0.2]. \\ Q(a_1, a_3) = p_1 + \frac{1}{2}p_2 \in [0.2, 0.5]. \\ Q(a_2, a_3) = p_1 + p_2 \in [0.3, 0.6].$$

We deduce that, using statistical preference as our basic binary relation:

- $a_2 \succ_s^{\mathcal{P}} a_1$ and $a_2 \succ_w^{\mathcal{P}} a_1$.
- $a_3 \succ_s^{\mathcal{P}} a_1$ and $a_3 \equiv_w^{\mathcal{P}} a_1$.
- $a_2 \equiv_w^{\mathcal{P}} a_3$ and they are incomparable with respect to strong \mathcal{P} -preference.

Consequently, with respect to the strong preference, the car company is preferred to the computer company, while the car and the fast food company are incomparable. With respect to weak preference, the car company is also preferred to the computer company, while the fast food company is indifferent to the car and the computer companies. \blacklozenge

A decision problem with uncertain rewards

Assume next that we have precise information about the probabilities of the different states of nature but that we have imprecise information about the utilities associated with the different rewards. Let us model this case by means of a random set, as we discussed in Section 4.2.1.

Example 4.133 (Cont) Assume that the probability of the different states of nature is given by:

$$P(\theta_1) = 0.2 \quad P(\theta_2) = 0.25 \quad P(\theta_3) = 0.3 \quad P(\theta_4) = 0.25,$$

but that we cannot determine precisely the consequences associated with each combination (alternative, state). We model the available information by means of a set of possible consequences, that we summarise in the following table:

	θ_1	θ_2	θ_3	θ_4
a_1	$[s_4, s_5]$	$\{s_3\}$	$[s_2, s_3]$	$\{s_2\}$
a_2	$\{s_5\}$	$[s_3, s_4]$	$[s_3, s_5]$	$[s_2, s_4]$
a_3	$\{s_2\}$	$[s_3]$	$[s_3, s_5]$	$[s_3, s_4]$

Since again we have qualitative rewards, we shall use statistical preference to compare the different alternatives. Taking into account that the utility space is finite, we deduce from Proposition 4.78 that the comparison of the random sets associated with each of the alternatives reduces to the comparison of their maxima and minima measurable selections. Moreover, since the utility space is finite, $\succeq_{\text{SP}_2} \Leftrightarrow \succeq_{\text{SP}_3}$ and $\succeq_{\text{SP}_5} \Leftrightarrow \succeq_{\text{SP}_6}$.

Let us compare alternatives a_1, a_2 :

$$\begin{aligned} Q(\min a_1, \max a_2) &= 0. \\ Q(\min a_1, \min a_2) &= 0.25. \\ Q(\max a_1, \max a_2) &= 0.1. \\ Q(\max a_1, \min a_2) &= 0.5. \end{aligned}$$

Using Proposition 4.78, we conclude that $a_2 \succ_{\text{SP}_i} a_1$ for $i = 1, 2, 3, 5, 6$ and $a_1 \equiv_{\text{SP}_4} a_2$.

With respect to alternatives a_1 and a_3 , we obtain that:

$$\begin{aligned} Q(\min a_1, \max a_3) &= 0.325 \\ Q(\min a_1, \min a_3) &= 0.325. \\ Q(\max a_1, \max a_3) &= 0.325. \\ Q(\max a_1, \min a_3) &= 0.475. \end{aligned}$$

Using Proposition 4.78, we conclude that $a_3 \succ_{\text{SP}_i} a_1$ for $i = 4$ and as a consequence also for $i = 1, 2, 3, 5, 6$.

Finally, if we compare alternatives a_2 and a_3 , we obtain that:

$$\begin{aligned} Q(\min a_2, \max a_3) &= 0.325 \\ Q(\min a_2, \min a_3) &= 0.475. \\ Q(\max a_2, \max a_3) &= 0.725. \\ Q(\max a_2, \min a_3) &= 1. \end{aligned}$$

Using Proposition 4.78, we conclude that $a_2 \succ_{\text{SP}_i} a_3$ for $i = 2, 3$, $a_3 \succ_{\text{SP}_i} a_2$ for $i = 5, 6$, $a_2 \equiv_{\text{SP}_4} a_3$ and they are incomparable with respect to \succeq_{SP_1} . Hence, in this case the choice between a_2 and a_3 would depend on our attitude towards risk, which would determine if we focus on the best or the worst-case scenarios. Consequently, both the car and the fast food companies are preferred to the computer one. However, the preference between the car and fast food companies depends on the chosen criteria. \blacklozenge

A decision problem between sets of alternatives

Assume now that we have precise beliefs and utilities but the choice must be made between sets of alternatives instead of pairs. In that case, we shall apply the conditions and results from Section 4.1.

Example 4.133 (Cont) Assume now that we may invest our money in another company a_4 in the telecommunications area, and that the choice must be made between two portfolios: one –that we shall denote \mathcal{X} –made by alternatives a_1, a_2 , and another –denoted by \mathcal{Y} –made by a_3, a_4 . Assume that the rewards associated with each alternative are given by the following table:

	θ_1	θ_2	θ_3	θ_4
	0.2	0.25	0.3	0.25
a_1	75	60	55	50
a_2	80	65	55	40
a_3	60	55	50	55
a_4	80	55	40	65

where the utilities are now expressed in a $[0, 100]$ scale.

If we compare these alternatives by means of stochastic dominance, we obtain that $a_1 \succ_{\text{FSD}} a_3$, $a_2 \succ_{\text{FSD}} a_4$ and any other pair (a_i, a_j) with $i \in \{1, 2\}, j \in \{3, 4\}$ are incomparable with respect to stochastic dominance. Hence, $\mathcal{X} \succ_{\text{FSD}_i} \mathcal{Y}$ for $i = 3, 4, 6$ and they are incomparable with respect to \succeq_{FSD_i} for $i = 1, 2, 5$.

Note that this example is an instance where \succeq_{FSD_2} is not equivalent to \succeq_{FSD_3} and \succeq_{FSD_5} is not equivalent to \succeq_{FSD_6} , because there is neither a maximum nor a minimum in the sets of distribution functions associated with \mathcal{X}, \mathcal{Y} .

On the other hand, if we compare any two alternatives by means of statistical preference, we obtain the following profile of preferences:

$$\mathbb{Q}_{\mathcal{X}, \mathcal{Y}} := \begin{pmatrix} 0.75 & 0.55 \\ 0.75 & 0.65 \end{pmatrix}.$$

Using Remark 4.73, we obtain that $\mathcal{X} \succ_{\text{SP}_1} \mathcal{Y}$, and as a consequence $\mathcal{X} \succ_{\text{SP}_i} \mathcal{Y}$ for $i = 2, \dots, 6$ and also $\mathcal{X} \succ_{\text{SP}_{\text{mean}}} \mathcal{Y}$. Hence, from the point of view of statistical preference the first portfolio should be preferred to the second. \blacklozenge

4.5 Conclusions

In this chapter we have considered the comparison of alternatives under both uncertainty and imprecision. As in Chapter 3, alternatives defined under uncertainty have been modelled by means of random variables, while the imprecision about the random variables has been modelled with sets of random variables, or in a more general situation, imprecise probability models.

We have extended binary relations to the comparison of sets of random variables instead of pairs of them. For this aim, we considered six possible generalisations. We have seen that the interpretation of each extension is related to the extensions of expected utility within imprecise probabilities.

We have mainly focused on two stochastic orders in this report: stochastic dominance and statistical preference. When we consider the binary relation to be first degree stochastic dominance, its extensions are related to the comparison of the p-boxes associated with the sets of random variables to compare. Also, according to the usual characterisation of stochastic dominance in terms of the comparison of the expectation of the increasing transformations of the random variables, we can also relate imprecise stochastic dominance to the comparison of the upper or lower expectations of the increasing transformation of the sets of random variables. We have also seen that our approach to extend stochastic dominance to the comparison of sets of random variables includes Denoeux approach ([61]) as a particular case, and we have also applied stochastic dominance to the comparison of possibility measures.

The extension of statistical preference has been connected to the comparison of the lower and upper medians of some set of random variables. We have seen that, when the sets of random variables to compare are finite, their comparison can be made by means of the pointwise comparison of the random variables by means of statistical preference, aggregating them with an aggregation function, and we have showed that the six extensions of statistical preference can be expressed in terms of aggregation functions.

We have also investigated two situations which can be considered as particular cases of the comparison of sets of random variables. On the one hand, we considered two random variables with imprecision on the utilities. That is, imprecise knowledge about the value of $X(\omega)$ and $Y(\omega)$. To model this imprecision, we have considered random sets Γ_X and Γ_Y , with the interpretation that the real value of $X(\omega)$ (respectively, $Y(\omega)$) belongs to $\Gamma_X(\omega)$ (respectively $\Gamma_Y(\omega)$). Then, we know that the random variables X, Y to be compared belong to the set of measurable selections of the random sets. Thus, the comparison of the random variables with imprecise utilities is made by the comparison of the random sets, which in fact can be made by means of the comparison of their associated sets of measurable selections.

On the other hand, we have also considered two random variables defined in a probability space whose probability is imprecisely described. We modelled this lack of information by means of a credal set. Then the random variables depend on the exact probability of the initial space. To deal with this imprecision we have introduced two new definitions: strong and weak preference.

We have seen that some binary relations, such as statistical preference, depend on the joint distribution of the random variables. In this framework Sklar's Theorem is a powerful tool that allows to build the joint distribution function from the marginals. However, there could be imprecision either in the marginal distributions, for example by considering p-boxes instead of distribution functions, or in the copula that links these marginals. For this reason we have developed a mathematical model that allows us to deal with this problem. In the first step, we showed that the infimum and supremum of sets of bivariate distribution functions are not bivariate distribution functions in general, because it may not satisfy the rectangle inequality. We have studied this problem by

means of imprecise probabilities, extending the notion of p-box to the bivariate case. Then, the infimum and supremum of bivariate distribution functions determine a coherent lower probability that satisfies some imprecise version of the rectangle inequalities.

On the other hand we have considered the case where the lack of information lies in the copula that links the marginals. For this problem, we have extended copulas to the imprecise framework, and we have proven an imprecise version of the Sklar's Theorem. Finally, we have seen how bivariate p-boxes and this imprecise version of the Sklar's Theorem could be applied to one and two-dimensional stochastic orders.

Since in the real life it is common to encounter situations in which the information is imprecisely described, the results of this chapter have several applications. We have showed how imprecise stochastic dominance can be applied in the comparison of Lorenz Curves and cancer survival rates, and illustrated the usefulness of imprecise statistical preference for multicriteria decision making problems under uncertainty.

5 Comparison of alternatives under imprecision

Chapter 3 was devoted to the comparison of alternatives in a decision problem under a context of uncertainty, where these alternatives were modelled by means of random variables. In Chapter 4 we added imprecision to the original problem, and we studied the comparison of sets of random variables. In this chapter we shall assume that the alternatives are defined under imprecision but without uncertainty. In this case we need not use probability theory, as the outcomes of the different alternative will be constant. However, the imprecision makes crisp sets not to an adequate model of the available information. Because of this, we shall use a more flexible theory than the one of crisp sets: that of fuzzy sets or any of its extensions, such as the theory of IF-sets or IVF-sets.

While for the comparison of random variables or sets of random variables we use stochastic orders, and some tools of the imprecise probability theory, for the comparison of IF-sets or IVF-sets we shall use some measures of comparison of these kinds of sets.

In the framework of fuzzy set theory, we can find in the literature several measures of comparison between fuzzy sets. The more usual measures of comparison are dissimilarities ([119]), dissimilitudes ([44]) and divergences ([159]), in addition to classical distances. Other authors, like Bouchon-Meunier ([27]) tried to define a general measure of comparison between fuzzy sets, that include the cited measures as particular cases. The last attempt was made by Couso et al. ([45]) where some usual axioms required by the measures of comparison of fuzzy sets are collected and analyzed.

Montes ([159]) made a complete study of divergences as comparison measures of fuzzy sets. She introduced a particular kind of divergences, the so-called local divergences, which have been proved to be very useful.

However, in the framework of IF-sets, in the literature we can only find distances for IF-sets and a lot of examples of IF-dissimilarities (see for example [36, 37, 85, 89, 92, 111, 113, 114, 138, 193, 212]), but there is not a thorough mathematical theory of comparison of IF-sets.

For this reason, the first part of this chapter is devoted to the generalization of the comparison measures from fuzzy sets to IF-sets. Note that even though in this part we shall deal with IF-sets, our comments in Section 2.3 guarantee that all our results remain valid for IVF-sets.

Afterwards, we shall investigate the relationship between IF-sets and imprecise probabilities. In this second part, we shall interpret IF-sets as IVF-sets, because this allows for a clearer connection to imprecise probability. Thus, we shall assume that the IVF-set is defined on a probability space, and that it may be thus interpreted as a random set. Then, we shall investigate its main properties.

The results we present in this chapter have several applications. On the one hand, the measures of comparison of IF-sets have been used in several fields, such as pattern recognition ([92, 93, 94, 113, 114]) or decision making ([194, 211]), among others. On the other hand, the connection between IVF-sets and imprecise probabilities will be very useful when producing a graded version of stochastic dominance, and they shall allow us to propose a generalization of stochastic dominance that allow the comparison of more than two sets of cumulative distribution functions.

5.1 Measures of comparison of IF-sets

In this section we are going to introduce some comparison measures for IF-sets. We begin by recalling the most common comparison measures for IF-sets: distances and dissimilarities.

Definition 5.1 A map $d : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$ is a distance between IF-sets if it satisfies the following properties:

- Positivity: $d(A, B) \geq 0$ for every $A, B \in IFSs(\Omega)$.
- Identity of indiscernibles: $d(A, B) = 0$ if and only if $A = B$.
- Symmetry: $d(A, B) = d(B, A)$ for every A and B in $IFSs(\Omega)$.
- Triangle inequality: $d(A, C) \leq d(A, B) + d(B, C)$ for every $A, B, C \in IFSs(\Omega)$.

Definition 5.2 A map $D : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$ is a dissimilarity for IF-sets (IF-dissimilarity, for short) if it satisfies the following axioms:

- IF-Diss.1: $D(A, A) = 0$ for every $A \in IFSs(\Omega)$.
- IF-Diss.2: $D(A, B) = D(B, A)$ for every $A, B \in IFSs(\Omega)$.
- IF-Diss.3: For every $A, B, C \in IFSs(\Omega)$ such that $A \subseteq B \subseteq C$ it holds that $D(A, C) \geq \max(D(A, B), D(B, C))$.

Remark 5.3 Some authors (see for instance [93, 113, 211]) replace axiom IF-Diss.1 by a stronger condition:

$$\widehat{\text{IF-Diss.1}}: \quad D(A, B) = 0 \Leftrightarrow A = B.$$

Thus, an IF-dissimilarity that satisfies $\widehat{\text{IF-Diss.1}}$ is more restrictive than IF-dissimilarities. Here, we shall restrict ourselves to the usual definition of IF-dissimilarity because it is more common in the literature. \blacklozenge

There are several examples of dissimilarities in the literature, as we shall see in Section 5.1.3. However, since its definition is not too restrictive, it is possible to define a counterintuitive measure of comparison for which axioms IF-Diss.1, IF-Diss.2 and IF-Diss.3 hold. In order to overcome this problem, we propose a measure of comparison of IF-sets called IF-divergence that satisfies the following natural properties:

- The divergence between two IF-sets is positive.
- The divergence between an IF-set and itself must be zero.
- The divergence between two IF-sets A and B is the same than the divergence between B and A . That is, it must be a symmetric function.
- The “more similar” two IF-sets are, the smaller is the divergence between them.

This is formally defined as follows.

Definition 5.4 Let us consider a function $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$. It is a divergence for IF-sets (IF-divergence for short) when it satisfies the following axioms:

- IF-Diss.1: $D_{\text{IFS}}(A, A) = 0$ for every $A \in \text{IFSS}(\Omega)$.
- IF-Diss.2: $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A)$ for every $A, B \in \text{IFSS}(\Omega)$.
- IF-Div.3: $D_{\text{IFS}}(A \cap C, B \cap C) \leq D_{\text{IFS}}(A, B)$, for every $A, B, C \in \text{IFSS}(\Omega)$.
- IF-Div.4: $D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B)$, for every $A, B, C \in \text{IFSS}(\Omega)$.

Note that IF-divergences are more restrictive than IF-dissimilarities. In order to prove this, let us first give a preliminary result.

Lemma 5.5 Let D_{IFS} be an IF-divergence, and let A, B, C and D be IF-sets such that $A \subseteq C \subseteq D \subseteq B$. Then $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(C, D)$.

Proof: Note that, if N and M are two IF-sets such that $N \subseteq M$, then $N \cup M = M$ and $N \cap M = N$. Then, it holds that:

$$\begin{aligned} C \subseteq D &\Rightarrow C \cap D = C, & D \subseteq B &\Rightarrow B \cap D = D, \\ A \subseteq C &\Rightarrow C \cup A = C, & C \subseteq B &\Rightarrow B \cup C = B. \end{aligned}$$

Using axioms IF-Div.3 and IF-Div.4 we obtain that:

$$\begin{aligned} D_{\text{IFS}}(C, D) &= D_{\text{IFS}}(C \cap D, B \cap D) \leq D_{\text{IFS}}(C, B) \\ &= D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B). \end{aligned}$$

We conclude that $D_{\text{IFS}}(C, D) \leq D_{\text{IFS}}(A, B)$. ■

Using this lemma we can prove now that every IF-divergence is also an IF-dissimilarity.

Proposition 5.6 *Every IF-divergence is an IF-dissimilarity.*

Proof: Let D_{IFS} be an IF-divergence, and let us check that it is also an IF-dissimilarity. For this, it suffices to prove that it satisfies axiom IF-Diss.3, because first and second axioms of IF-divergences and IF-dissimilarities coincide. Let A, B and C be three IF-sets such that $A \subseteq B \subseteq C$. Then, taking into account that $A \subseteq A \subseteq B \subseteq C$, and applying the previous lemma, $D_{\text{IFS}}(A, C) \geq D_{\text{IFS}}(A, B)$. On the other hand, since $A \subseteq B \subseteq C \subseteq C$, the previous lemma also implies that $D_{\text{IFS}}(A, C) \geq D_{\text{IFS}}(B, C)$.

Hence, D_{IFS} satisfies axiom Diss.3 and, consequently, it is a dissimilarity. ■

We have seen that every IF-divergence is also an IF-dissimilarity. In Example 5.8 we will see that the converse does not hold in general.

In the fuzzy framework Couso et al. ([44]) introduced a measure of comparison called dissimilitude. It can be generalized to the comparison of IF-sets in the following way.

Definition 5.7 *A map $D : \text{IFSs}(\Omega) \times \text{IFSs}(\Omega) \rightarrow \mathbb{R}$ is an IF-dissimilitude if it satisfies the following properties:*

- IF-Diss.1:* $D_{\text{IFS}}(A, A) = 0$ for every $A \in \text{IFSs}(\Omega)$.
- IF-Diss.2:* $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A)$ for every $A, B \in \text{IFSs}(\Omega)$.
- IF-Diss.3:* If $A, B, C \in \text{IFSs}(\Omega)$ satisfies $A \subseteq B \subseteq C$, then $D_{\text{IFS}}(A, C) \geq \max(D_{\text{IFS}}(A, B), D_{\text{IFS}}(B, C))$.
- IF-Div.4:* $D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B)$, for every $A, B, C \in \text{IFSs}(\Omega)$.

This measure of comparison is stronger than IF-dissimilarities, but less restrictive than IF-divergences. Moreover, the converse implications do not hold in general. Let us give an example of an IF-dissimilitude that is not an IF-divergence and an example of an IF-dissimilarity that is not an IF-dissimilitude.

Example 5.8 *First of all, we are going to build a dissimilarity that is not a dissimilitude.*

Let us consider the function $D : \text{IFSs}(\Omega) \times \text{IFSs}(\Omega) \rightarrow [0, 1]$ defined on a finite Ω by:

$$D(A, B) = \left| \max_{\omega \in \Omega} (\max(0, \mu_B(\omega) - \mu_A(\omega))) - \max_{\omega \in \Omega} (\max(0, \mu_A(\omega) - \mu_B(\omega))) \right|.$$

Let us see that D is an IF-dissimilarity:

IF-Diss.1: $D(A, A) = 0$, since $\mu_B(\omega) - \mu_A(\omega) = 0$ for any $\omega \in \Omega$.

IF-Diss.2: Obviously, $D(A, B) = D(B, A)$.

IF-Diss.3: Let A, B and C be three IF-sets such that $A \subseteq B \subseteq C$. Then, since $\mu_A(\omega) \leq \mu_B(\omega) \leq \mu_C(\omega)$, it holds that:

$$\begin{aligned} D(A, B) &= |\max_{\omega \in \Omega} \mu_B(\omega) - \mu_A(\omega)|, \\ D(B, C) &= |\max_{\omega \in \Omega} \mu_C(\omega) - \mu_B(\omega)|, \\ D(A, C) &= |\max_{\omega \in \Omega} \mu_C(\omega) - \mu_A(\omega)|. \end{aligned}$$

Moreover,

$$\mu_C(\omega) - \mu_A(\omega) \geq \max(\mu_C(\omega) - \mu_B(\omega), \mu_B(\omega) - \mu_A(\omega)),$$

and therefore:

$$D(A, C) \geq \max(D(A, B), D(B, C)).$$

Thus, D satisfies axiom IF-Diss.3 and therefore it is an IF-dissimilarity. Let us show that D is not a dissimilitude, or equivalently, that there are IF-sets A, B and C such that $D(A \cup C, B \cup C) > D(A, B)$. To see this, let us consider $\Omega = \{\omega_1, \omega_2\}$ and define the IF-sets A and B by:

$$\begin{aligned} A &= \{(\omega_1, 0.5, 0), (\omega_2, 0, 0)\}, & B &= \{(\omega_1, 0, 0), (\omega_2, 0.6, 0)\}, \\ C &= \{(\omega_1, 0.5, 0), (\omega_2, 0.2, 0)\}. \end{aligned}$$

It holds that:

$$\begin{aligned} A \cup C &= \{(\omega_1, 0.5, 0), (\omega_2, 0.2, 0)\}, \\ B \cup C &= \{(\omega_1, 0.5, 0), (\omega_2, 0.6, 0)\}. \end{aligned}$$

Then:

$$D(A, B) = |0.5 - 0.6| = 0.1 \not\geq 0.4 = |0.2 - 0.6| = D(A \cup C, B \cup C).$$

Hence, D does not fulfill Div.4, and therefore it is neither an IF-dissimilitude nor an IF-divergence. \blacklozenge

Example 5.9 Let us give an IF-dissimilitude that is not an IF-divergence. Consider the function D defined by:

$$D(A, B) = \begin{cases} 1 & \text{if } A = \emptyset \text{ or } B = \emptyset, \text{ but } A \neq B. \\ 0 & \text{otherwise.} \end{cases}$$

Let us see that this function is a dissimilitude:

IF-Diss.1: $D(A, A) = 0$ by definition.

IF-Diss.2: D is symmetric by definition.

IF-Diss.3: Let A , B and C be three IF-sets such that $A \subseteq B \subseteq C$. Then,

$$\mu_A(\omega) \leq \mu_B(\omega) \leq \mu_C(\omega) \text{ and } \nu_A(\omega) \geq \nu_B(\omega) \geq \nu_C(\omega)$$

for every $\omega \in \Omega$.

There are two cases: on the one hand, if $D(A, C) = 1$, then

$$D(A, C) = 1 \geq \max(D(A, B), D(B, C)).$$

On the other hand, $A \neq \emptyset$ and $C \neq \emptyset$ or $A = C$. Since $A \subseteq B \subseteq C$, in the first case $B \neq \emptyset$ and in the second one $B = A = C$. In all cases, $D(A, C) = D(A, B) = D(B, C) = 0$.

Div.4: Let us show that $D(A \cup C, B \cup C) \leq D(A, B)$ for every IF-sets A, B and C . This inequality holds if $D(A, B) = 1$. Otherwise, if $D(A, B) = 0$ then $A \neq \emptyset$ and $B \neq \emptyset$ or $A = B$. Since $A \subseteq A \cup C$ and $B \subseteq B \cup C$, in the first case we deduce that $A \cup C \neq \emptyset$ and $B \cup C \neq \emptyset$ and we conclude that $D(A \cup C, B \cup C) = D(A, B) = 0$. In the second case, $D(A \cup C, B \cup C) = D(A \cup C, A \cup C) = 0 = D(A, B)$.

Thus, D is an IF-dissimilitude, but it is not an IF-divergence since it does not fulfill axiom Div.3: if we consider the IF-sets A, B and C defined by

$$\begin{aligned} A &= \{(\omega_0, 0, 1), (\omega, \mu_A(\omega), \nu_A(\omega)) \mid \omega \neq \omega_0\}; \\ B &= \{(\omega, \mu_B(\omega), \nu_B(\omega)) \mid \omega \in \Omega\}; \\ C &= \{(\omega_0, 1, 0), (\omega, 0, 1) \mid \omega \neq \omega_0\}; \end{aligned}$$

where $\mu_B(\omega) > 0$ for every $\omega \in \Omega$ and $\mu_A(\omega) = \mu_B(\omega)$ for every $\omega \neq \omega_0$, for a fixed element ω_0 of Ω ; then, $A \cap C = \emptyset$ but $B \cap C \neq \emptyset$, and therefore:

$$D(A \cap C, B \cap C) = 1 > 0 = D(A, B).$$

Hence, D is an IF-dissimilitude that is not an IF-divergence.

We have already studied the relationships among IF-dissimilarities, IF-divergences and IF-dissimilitudes, and we have also mentioned some counterexamples related to the distance. In fact, that there is not a general relationship between the notion of distance for IF-sets and these three measures of comparison. To show that, we start with an example of an IF-distance that is not an IF-dissimilarity.

Example 5.10 Let us consider the function D defined by:

$$D(A, B) = \begin{cases} 0 & \text{if } A = B, \\ \frac{1}{2} & \text{if } A - B \neq \emptyset \text{ or } B - A \neq \emptyset \text{ and } \mu_{A \cap B}(\omega) = 0.3 \forall \omega \in \Omega, \\ 1 & \text{otherwise,} \end{cases}$$

where the IF-difference is the one of Example 2.56. Let us see that this function is a distance for IF-sets.

Positivity: By definition, $D(A, B) \geq 0$ for every $A, B \in \text{IFSSs}(\Omega)$.

Identity of indiscernibles: By definition, $D(A, B) = 0$ if and only if $A = B$.

Symmetry: D is also symmetric by definition.

Triangular inequality: Let us see that $D(A, C) \leq D(A, B) + D(B, C)$ holds for any $A, B, C \in \text{IFSSs}(\Omega)$. On the one hand, if $D(A, C) = 0$, the inequality trivially holds. If $D(A, B) = \frac{1}{2}$, we can assume, without loss of generality, that $A - C \neq \emptyset$, and then, $A \neq C$. This implies that either $A \neq B$ or $B \neq C$, and consequently either $D(A, B) \geq \frac{1}{2}$ or $D(B, C) \geq \frac{1}{2}$. Therefore the inequality holds. Finally, if $D(A, C) = 1$ and we assume that the triangle inequality does not hold, then without loss of generality we can assume that $D(A, B) = 0$. In that case, $A = B$, and therefore $D(A, C) = D(B, C) = 1$, a contradiction arises. We conclude that the triangle inequality holds.

Thus, D is a distance for IF-sets. However, it is not an IF-dissimilarity, since we can find IF-sets A, B and C , with $A \subseteq B \subseteq C$, such that $D(A, C) < D(A, B)$: let us consider $\Omega = \{\omega\}$ and the IF-sets A, B and C defined by:

$$A = \{(\omega, 0.2, 0.4)\}, \quad B = \{(\omega, 0.3, 0.2)\}, \quad C = \{(\omega, 0.4, 0)\}.$$

It is obvious that $A \subseteq B \subseteq C$. Moreover, it holds that:

$$D(A, C) = 1 \text{ and } D(B, C) = 0.5.$$

We conclude that D is not an IF-dissimilarity. \blacklozenge

We have seen that IF-distances are not IF-dissimilarities in general. Thus, they cannot be, in general, IF-divergences or IF-dissimilitudes, since in that case they would be in particular IF-dissimilarities. We next show that the converse implications do not hold either.

Example 5.11 Let us give an example of an IF-divergence that is not a distance between IF-sets. Consider the function D defined by:

$$D(A, B) = \max_{\omega \in \Omega} (\max(0, \mu_A(\omega) - \mu_B(\omega)))^2 + \max_{\omega \in \Omega} (\max(0, \mu_A(\omega) - \mu_B(\omega)))^2.$$

IF-Div.1: It is obvious that $D(A, A) = 0$.

IF-Div.2: By definition, D is also symmetric.

IF-Div.3: Let us prove that $D(A, B) \geq D(A \cap C, B \cap C)$ for any A, B, C . Using the first part of Lemma A.1 in Appendix A, for any ω it holds that:

$$\max(0, \mu_A(\omega) - \mu_B(\omega)) \geq \max(0, \min(\mu_A(\omega), \mu_C(\omega)) - \min(\mu_B(\omega), \mu_C(\omega))).$$

It trivially follows that $D(A, B) \geq D(A \cap C, B \cap C)$.

IF-Div.4: Similarly, let us prove that $D(A, B) \geq D(A \cup C, B \cup C)$ for any A, B, C . Taking into account again the first part of Example A.1 in Lemma A, any ω satisfies the following:

$$\max(0, \mu_A(\omega) - \mu_B(\omega)) \geq \max(0, \max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))).$$

This implies that $D(A, B) \geq D(A \cup C, B \cup C)$.

We conclude that D is an IF-divergence. However, it does not satisfy the triangular inequality, because for the IF-sets A , B and C of $\Omega = \{\omega\}$, defined by:

$$A = \{(\omega, 0, 1)\}, \quad B = \{(\omega, 0.4, 0)\} \text{ and } C = \{(\omega, 0.5, 0)\},$$

it holds that:

$$D(A, C) = 0.25 \not\leq 0.16 + 0.01 = D(A, B) + D(B, C).$$

Thus, D does not satisfy the triangular inequality. ♦

Since the measure defined in this example is an IF-divergence, it is also an IF-dissimilarity and an IF-dissimilitude. Then, we can see that none of these measures satisfy, in general, the properties that define a distance.

Let us show next that an IF-dissimilitude and a distance is not necessarily an IF-divergence.

Example 5.12 Let us consider the map

$$D : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$$

defined by:

$$D(A, B) = \begin{cases} 0 & \text{if } A = B. \\ 1 & \text{if } A \neq B \text{ and either } \mu_A(\omega) = 0 \forall \omega \in \Omega \text{ or } \mu_B(\omega) = 0 \forall \omega \in \Omega. \\ 0.5 & \text{otherwise.} \end{cases}$$

First of all, let us prove that D is a distance for IF-sets.

Positivity: By definition, $D(A, B) \geq 0$ for every $A, B \in IFSs(\Omega)$.

Identity of indiscernibles: By definition, $D(A, B) = 0$ if and only if $A = B$.

Triangular inequality: Let us consider $A, B, C \in IFSs(\Omega)$, and let us prove that $D(A, C) \leq D(A, B) + D(B, C)$. If $D(A, C) = 0$, obviously the inequality holds. If $D(A, C) = 0.5$, then $A \neq C$, and therefore either $A \neq B$, and consequently $D(A, B) \geq 0.5$ or $B \neq C$, and consequently $D(B, C) \geq 0.5$. Then, $D(A, B) + D(B, C) \geq 0.5 = D(A, C)$.

Otherwise, $D(A, C) = 1$. In such a case, $A \neq C$ and we can assume that $\mu_A(\omega) = 0$ for every $\omega \in \Omega$. Then, if $A \neq B$, $D(A, B) = 1$, and if $A = B$, then $D(B, C) = D(A, C) = 1$. We conclude thus that the triangular inequality holds.

Let us now prove that D is also an IF-dissimilitude:

IF-Diss.1: We have already seen that $D(A, A) = 0$.

IF-Diss.2: Obviously, D is symmetric.

IF-Diss.3: Consider $A, B, C \in \text{IFSS}(\Omega)$ such that $A \subseteq B \subseteq C$, and let us prove that $D(A, C) \geq \max(D(A, B), D(B, C))$. Note that if $D(A, C) = 0$, then $A = B = C$, and therefore the inequality holds. Moreover, if $D(A, C) = 1$ then the inequality also holds because $\max(D(A, B), D(B, C)) \leq 1$. Finally, assume that $D(A, C) = 0.5$. In such a case $A \neq C$, and therefore either $A \neq B$ or $B \neq C$, and there is $\omega \in \Omega$ such that $\mu_C(\omega) \geq \mu_A(\omega) > 0$. Then, as $\mu_C(\omega) \geq \mu_B(\omega) \geq \mu_A(\omega)$, $D(A, B), D(B, C) \leq 0.5$. Thus, axiom IF-Diss.3 holds.

IF-Div.4: Let us now consider three IF-sets A, B and C , and let us prove that $D(A \cup C, B \cup C) \leq D(A, B)$. First of all, if $D(A, B) = 1$, then the previous inequality trivially holds, since D is bounded by 1. Moreover, if $D(A, B) = 0$ then $A = B$, and consequently applying IF-Diss.1 $D(A \cup C, B \cup C) = D(A \cup C, A \cup C) = 0$. Finally, assume that $D(A, B) = 0.5$. In such a case, $A \neq B$ and there exist $\omega_1, \omega_2 \in \Omega$ such that $\mu_A(\omega_1) > 0$ and $\mu_B(\omega_2) > 0$. Let us note that:

$$\begin{aligned}\mu_{A \cup C}(\omega) &= \max(\mu_A(\omega), \mu_C(\omega)) \geq \mu_A(\omega) \text{ and} \\ \mu_{B \cup C}(\omega) &= \max(\mu_B(\omega), \mu_C(\omega)) \geq \mu_B(\omega).\end{aligned}$$

Consequently, $\mu_{A \cup C}(\omega_1) \geq \mu_A(\omega_1) > 0$ and $\mu_{B \cup C}(\omega_2) \geq \mu_B(\omega_2) > 0$. Then it holds that $D(A \cup C, B \cup C) \leq 0.5 = D(A, B)$.

Thus, D is a distance and an IF-dissimilitude. Let us show that it is not an IF-divergence. Consider $\Omega = \{\omega_1, \omega_2\}$ and the IF-sets A, B and C defined by:

$$\begin{aligned}A &= \{(\omega_1, 1, 0), (\omega_2, 0, 0)\}. \\ B &= \{(\omega_1, 1, 0), (\omega_2, 1, 0)\}. \\ C &= \{(\omega_1, 0, 0), (\omega_2, 1, 0)\}.\end{aligned}$$

Then:

$$\begin{aligned}A \cap C &= \{(\omega_1, 0, 0), (\omega_2, 0, 0)\}. \\ B \cap C &= \{(\omega_1, 0, 0), (\omega_2, 1, 0)\}.\end{aligned}$$

Then, $D(A, B) = 0.5$ and $D(A \cap C, B \cap C) = 1$, and therefore

$$D(A \cap C, B \cap C) > D(A, B),$$

a contradiction with IF-Div.3. Thus D cannot be an IF-divergence. \blacklozenge

To conclude this part, it only remains to show that if D is an IF-dissimilarity and a distance, it is not necessarily an IF-dissimilitude.

Example 5.13 Consider the map

$$D : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$$

defined by:

$$D(A, B) = \begin{cases} 0 & \text{if } A = B. \\ 1 & \text{if } A \neq B \text{ and either } A = \Omega \text{ or } B = \Omega. \\ 0.5 & \text{otherwise.} \end{cases}$$

Let us prove that D is a distance for IF-sets.

Positivity, the identity of indiscernibles and symmetry trivially hold. Let us prove that the triangular inequality is also satisfied. Let A, B and C be three IF-sets, and let us see that $D(A, C) \leq D(A, B) + D(B, C)$.

- If $D(A, C) = 0$, the inequality trivially holds.
- If $D(A, C) = 0.5$, then $A \neq C$, and therefore either $A \neq B$ or $B \neq C$, and consequently $D(A, B) + D(B, C) \geq 0.5 = D(A, C)$.
- Finally, if $D(A, C) = 1$, we can assume, without loss of generality, that $A = \Omega$. Then, if $B = A$, $D(B, C) = 1$, and therefore $D(A, C) = 1 = D(A, B) + D(B, C)$. Otherwise, if $B \neq A$, then $D(A, B) = 1$, and therefore

$$D(A, C) = 1 \leq D(A, B) + D(B, C).$$

Thus, D is a distance for IF-sets.

Let us now prove that it is also an IF-dissimilarity. On the one hand, properties IF-Diss.1 and IF-Diss.2 are trivially satisfied. Let us see that IF-Diss.3 also holds. Consider three IF-sets A, B, C satisfying $A \subseteq B \subseteq C$, and let us prove that $D(A, C) \geq \max(D(A, B), D(B, C))$.

- If $D(A, C) = 1$, obviously $D(A, C) \geq \max(D(A, B), D(B, C))$.
- If $D(A, C) = 0.5$, then $A \neq C$ and there is $\omega \in \Omega$ such that $\mu_A(\omega) \leq \mu_B(\omega) \leq \mu_C(\omega) < 1$. Then, $\max(D(A, B), D(B, C)) \leq 0.5 = D(A, C)$.
- Finally, if $D(A, C) = 0$, $A = B = C$ holds, and then $D(A, B) = D(B, C) = 0$.

Thus, D is a distance for IF-sets and an IF-dissimilarity. However, it is not an IF-dissimilitude, for it does not satisfy axiom IF-Div.4: to see this, consider the universe $\Omega = \{\omega_1, \omega_2\}$, and the IF-sets

$$A = \{(\omega_1, 1, 0), (\omega_2, 0, 0)\} \text{ and } B = \{(\omega_1, 0, 0), (\omega_2, 1, 0)\}.$$

It holds that $D(A, B) = 0.5$. However, if we consider $C = B$, then $A \cup C = \Omega$, and therefore:

$$D(A \cup C, B \cup C) = D(\Omega, B) = 1.$$

Then, $D(A \cup C, B \cup C) > D(A, B)$, and therefore axiom IF-Div.4 is not satisfied. This shows that D is not an IF-divergence. \blacklozenge

Figure 5.1 summarizes the relationships between the different methods for comparing IF-sets.

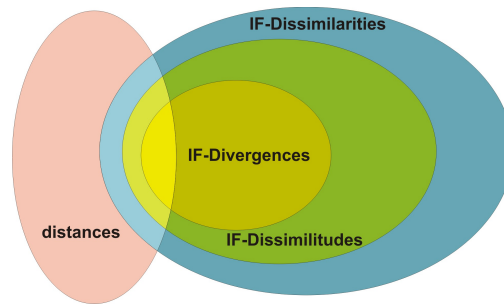


Figure 5.1: Relationships among IF-divergences, IF-dissimilitudes, IF-dissimilarities and distances for IF-sets.

5.1.1 Theoretical approach to the comparison of IF-sets

Bouchon-Meunier et al. ([27]) proposed a general measure of comparison for fuzzy sets that generates some particular measures depending on the conditions imposed to such a general measure.

Following this ideas, in this section we define a general measure of comparison between IF-sets that, depending on the imposed properties, generates either distances, or IF-dissimilarities or IF-divergences.

For this, let us consider a function $D : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$, and assume that there is a generator function G_D :

$$G_D : IFSs(\Omega) \times IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}^+ \quad (5.1)$$

such that D can be expressed by:

$$D(A, B) = G_D(A \cap B, B - A, A - B),$$

where $-$ is a difference operator for IF-sets, according to Definition 2.55, that fulfills D3, D4 and D5.

We shall see that depending on the conditions imposed on G_D , we can obtain that D is either an IF-dissimilarity, an IF-divergence or a distance for IF-sets.

We begin by determining which conditions must be imposed on G_D in order to obtain a distance for IF-sets.

Proposition 5.14 *Consider the function $D : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$ that can be expressed as in Equation (5.1) by means of a generator $G_D : IFSs(\Omega) \times IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}^+$. If the function G_D satisfies the properties:*

- S-Dist.1:* $G_D(A, B, C) = 0$ if and only if $B = C = \emptyset$;
- S-Dist.2:* $G_D(A, B, C) = G_D(A, C, B)$ for every $A, B, C \in IFSs(\Omega)$;
- S-Dist.3:* For every $A, B, C \in IFSs(\Omega)$,
 $G_D(A \cap C, C - A, A - C) \leq G_D(A \cap B, B - A, A - B)$
 $+ G_D(B \cap C, C - B, B - C)$;

then D is a distance for IF-sets.

Proof: Let us prove that D satisfies the axioms of IF-distances.

Positivity: it trivially follows from the positivity of G_D . To show the identity of indiscernibles, let A and B be two IF-sets. Then, by property S-Dist.1:

$$D(A, B) = G_D(A \cap B, B - A, A - B) = 0 \Leftrightarrow B - A = A - B = \emptyset,$$

and by properties D1 and D5 this is equivalent to $A = B$.

Symmetry: Let A and B be two IF-sets. Using S-Dist.2, we have that:

$$\begin{aligned} D(A, B) &= G_D(A \cap B, B - A, A - B) \\ &= G_D(A \cap B, A - B, B - A) = D(B, A). \end{aligned}$$

Triangular inequality: Let A , B and C be three IF-sets. By S-Dist.3, it holds that:

$$\begin{aligned} D(A, C) &= G_D(A \cap C, C - A, A - C) \\ &\leq G_D(A \cap B, B - A, A - B) + G_D(B \cap C, C - B, B - C) \\ &= D(A, B) + D(B, C). \end{aligned} \quad \blacksquare$$

Let us now consider IF-dissimilarities. We have proven the following result:

Proposition 5.15 *Let D be a map $D : IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}^+$ that can be expressed as in Equation (5.1) by means of the generator G_D , where $G_D : IFSS(\Omega) \times IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}^+$. Then, D is an IF-dissimilarity if G_D satisfies the following properties:*

- S-Diss.1:* $G_D(A, \emptyset, \emptyset) = 0$ for every $A \in IFSS(\Omega)$.
- S-Dist.2:* $G_D(A, B, C) = G_D(A, C, B)$ for every $A, B, C \in IFSS(\Omega)$.
- S-Diss.3:* $G_D(A, B, \emptyset)$ is increasing in B .
- S-Diss.4:* $G_D(A, B, \emptyset)$ is decreasing in A .

Proof: Let us prove that D is an IF-dissimilarity.

IF-Diss.1: Let A be an IF-set. By D1 and S-Diss.1 it holds that

$$D(A, A) = G_D(A \cap A, A - A, A - A) = G_D(A, \emptyset, \emptyset) = 0.$$

IF-Diss.2: Let A and B be two IF-sets. Then, by S-Dist.2, D is symmetric:

$$\begin{aligned} D(A, B) &= G_D(A \cap B, B - A, A - B) \\ &= G_D(A \cap B, A - B, B - A) = D(B, A). \end{aligned}$$

IF-Diss.3: Let A, B and C be three IF-sets such that $A \subseteq B \subseteq C$, and let us prove that $D(A, C) \geq \max(D(A, B), D(B, C))$. First of all, let us compute $D(A, C), D(A, B)$ and $D(B, C)$.

$$\begin{aligned} D(A, C) &= G_D(A \cap C, C - A, A - C) = G_D(A, C - A, \emptyset). \\ D(A, B) &= G_D(A \cap B, B - A, A - B) = G_D(A, B - A, \emptyset). \\ D(B, C) &= G_D(B \cap C, C - B, B - C) = G_D(B, C - B, \emptyset). \end{aligned}$$

On one hand, let us prove that $D(A, C) \geq D(A, B)$. By D2, it holds that $B - A \subseteq C - A$, and therefore, by S-Diss.3:

$$D(A, C) = G_D(A, C - A, \emptyset) \geq G_D(A, B - A, \emptyset) = D(A, B).$$

Let us prove next that $D(A, C) \geq D(B, C)$. By D4 it holds that $C - B \subseteq C - A$, and therefore:

$$\begin{aligned} D(A, C) = G_D(A, C - A, \emptyset) &\stackrel{S-Diss.4}{\geq} G_D(B, C - A, \emptyset) \\ &\stackrel{S-Diss.3}{\geq} G_D(B, C - B, \emptyset) = D(B, C). \end{aligned}$$

Thus, we conclude that D is an IF-dissimilarity. ■

Concerning IF-divergences, we have established the following:

Proposition 5.16 *Let D be a map $D : IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}$ generated by G_D as in Equation (5.1), where $G_D : IFSS(\Omega) \times IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}^+$. Then, D is an IF-divergence if G_D satisfies the following properties:*

- S-Diss.1:* $G_D(A, \emptyset, \emptyset) = 0$ for every $A, B \in IFSs(\Omega)$.
S-Dist.2: $G_D(A, B, C) = G_D(A, C, B)$ for every $A, B, C \in IFSs(\Omega)$.
S-Div.3: $G_D(A, B, C)$ is increasing in B and C .
S-Div.4: $G_D(A, B, C)$ is independent of A .

Note that axiom S-Div.4 is a very strong condition. We require it because IF-divergences focus on the difference between the IF-sets instead of the intersection.

Proof: Let us prove that D is an IF-divergence.

First and second axioms of IF-divergences and IF-dissimilarities coincide. Furthermore, as we proved in Proposition 5.15, they follow from S-Diss.1 and S-Dist.2.

IF-Div.3: Let A, B and C be three IF-sets. Since the IF-difference operator fulfills D3, then $(A \cap C) - (B \cap C) \subseteq A - B$ and $(B \cap C) - (A \cap C) \subseteq B - A$. Therefore, by S-Div.3 and S-Div.4:

$$\begin{aligned} D(A \cap C, B \cap C) &= G_D(A \cap B \cap C, (B \cap C) - (A \cap C), (A \cap C) - (B \cap C)) \\ &= G_D(A \cap B, (B \cap C) - (A \cap C), (A \cap C) - (B \cap C)) \\ &\leq G_D(A \cap B, B - A, A - B) = D(A, B). \end{aligned}$$

IF-Div.4: Consider the IF-sets A, B and C . As in the previous axiom, applying property D4 of the IF-difference $-$, we obtain that $(A \cup C) - (B \cup C) \subseteq A - B$ and $(B \cup C) - (A \cup C) \subseteq B - A$. As a consequence,

$$\begin{aligned} D(A \cup C, B \cup C) &= G_D((A \cup C) \cap (B \cup C), (B \cup C) - (A \cup C), (A \cup C) - (B \cup C)) \\ &\stackrel{S-Div.4}{=} G_D(A \cap B, (B \cup C) - (A \cup C), (A \cup C) - (B \cup C)) \\ &\stackrel{S-Div.3}{\leq} G_D(A \cap B, B - A, A - B) = D(A, B). \end{aligned}$$

We conclude that D is an IF-divergence. ■

In order to find sufficient conditions over G_D so as to build an IF-dissimilitude D , we need D to satisfy axioms IF-Diss.1, IF-Diss.2, IF-Diss.3 and IF-Div.4. As we have already mentioned, axioms IF-Diss.1 and IF-Diss.2 are implied by conditions:

- S-Diss.1: $G_D(A, \emptyset, \emptyset)$ for every $A, B \in IFSs(\Omega)$.
S-Dist.2: $G_D(A, B, C) = G_D(A, C, B)$ for every $A, B \in IFSs(\Omega)$.

In order to prove condition IF-Div.4, in Proposition 5.16 we required the following:

- S-Div.3: $G_D(A, B, C)$ is increasing in B and C .
S-Div.4: $G_D(A, B, C)$ is independent of A .

Moreover, it is trivial that these conditions imply S-Diss.3 and S-Diss.4, that also follow from axiom IF-Diss.3. Therefore, the conditions that need to be imposed on G_D

in order to obtain an IF-dissimilitude are the same that we have imposed in order to obtain an IF-divergence.

Let us give an example of a function G_D that generates an IF-dissimilarity but not an IF-divergence.

Example 5.17 Consider the function $G_D : IFSS(\Omega) \times IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}^+$ defined, for every $A, B, C \in IFSS(\Omega)$, by:

$$G_D(A, B, C) = |\max_{\omega \in \Omega} \mu_B(\omega) - \max_{\omega \in \Omega} \mu_C(\omega)|.$$

This function generates an IF-dissimilarity because it satisfies properties *S-Diss.i*, with $i = 1, 3, 4$ and *S-Dist.2*.

S-Diss.1: By definition, $G_D(A, \emptyset, \emptyset) = 0$, since $\mu_\emptyset(\omega) = 0$ for every $\omega \in \Omega$.

S-Dist.2: G_D is symmetric with respect its second and third components:

$$\begin{aligned} G_D(A, B, C) &= |\max_{\omega \in \Omega} \mu_B(\omega) - \max_{\omega \in \Omega} \mu_C(\omega)| \\ &= |\max_{\omega \in \Omega} \mu_C(\omega) - \max_{\omega \in \Omega} \mu_B(\omega)| = G_D(A, C, B). \end{aligned}$$

S-Diss.3: Let A, B and B' be three IF-sets such that $B \subseteq B'$. Then, $\mu_B(\omega) \leq \mu_{B'}(\omega)$ for every $\omega \in \Omega$. Then it holds that:

$$G_D(A, B, \emptyset) = \max_{\omega \in \Omega} \mu_B(\omega) \leq \max_{\omega \in \Omega} \mu_{B'}(\omega) = G_D(A, B', \emptyset).$$

Thus, $G_D(A, B, \emptyset)$ is increasing in B .

S-Diss.4: It is obvious that G_D does not depend on its first component, and therefore, it is in particular decreasing on A .

Hence, G_D satisfies the conditions of Proposition 5.15, and therefore the map D defined by:

$$D(A, B) = G_D(A \cap B, B - A, A - B), \text{ for every } A, B \in IFSS(\Omega)$$

is an IF-dissimilarity. However, in general G_D does not satisfy *S-Div.4*. To see this, it is enough to consider the IF-difference of Example 2.56. In that case, the function G_D generates the IF-dissimilarity of Example 5.8, which was showed not to satisfy condition *IF-Div.4*. Then, D is neither an IF-dissimilitude nor an IF-divergence. This implies that G_D does not fulfill *S-Div.4*, because otherwise D would be an IF-divergence. ♦

Let us see next an example of a function G_D that generates an IF-divergence that is not a distance for IF-sets.

Example 5.18 Consider the function $G_D : IFSs(\Omega) \times IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}^+$ defined by:

$$G_D(A, B, C) = \left(\max_{\omega \in \Omega} \mu_B(\omega) \right)^2 + \left(\max_{\omega \in \Omega} \mu_C(\omega) \right)^2,$$

for every $A, B, C \in IFSs(\Omega)$. This function generates an IF-divergence, since it trivially satisfies the conditions in Proposition 5.16. However, it does not generate a distance for IF-sets. To see it, consider the IF-difference defined in Example 2.56. Then, the IF-divergence that generates G_D with this IF-difference coincides with the one given in Example 5.11, where we proved that it was not a distance for IF-sets. ♦

Finally, let us give an example of a function G_D that generates a distance for fuzzy sets that is not an IF-dissimilarity, and therefore it is neither an IF-divergence nor an IF-dissimilitude.

Example 5.19 Consider the function

$$G_D : IFSs(\Omega) \times IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}^+$$

by:

$$G_D(A, B, C) = \begin{cases} 0 & \text{if } B = C = \emptyset, \\ 0.5 & \text{if } B \neq \emptyset \text{ or } C \neq \emptyset \text{ and } \mu_A(\omega) = 0.3 \text{ for all } \omega \in \Omega, \\ 1 & \text{otherwise.} \end{cases}$$

Let us prove that G_D satisfies conditions of Proposition 5.14.

S-Dist.1: By definition, $G_D(A, B, C) = 0$ if and only if $B = C = \emptyset$.

S-Dist.2: Obviously, $G_D(A, B, C) = G_D(A, C, B)$ for every $A, B, C \in IFSs(\Omega)$.

S-Dist.3: Let us consider $A, B, C \in IFSs(\Omega)$, and we want to prove that

$$G_D(A \cap C, C - A, A - C) \leq G_D(A \cap B, B - A, A - B) + G_D(C \cap B, B - C, C - B).$$

- If $G_D(A \cap C, C - A, A - C) = 0$, then the inequality trivially holds.
- Let us now assume that $G_D(A \cap C, C - A, A - C) = 0.5$. Thus, either $A - C \neq \emptyset$ or $C - A \neq \emptyset$ and $\mu_{A \cap C}(\omega) = 0.3$ for every $\omega \in \Omega$. Let us note that, as $A \neq C$, either $A \neq B$ or $B \neq C$. Equivalently, either $G_D(A \cap B, B - A, A - B) \geq 0.5$ or $G_D(C \cap B, B - C, C - B) \geq 0.5$. Then, in this case the inequality also holds.
- Finally, consider the case where $G_D(A \cap C, C - A, A - C) = 1$. Then, $A - C \neq \emptyset$ or $C - A \neq \emptyset$ and $\mu_{A \cap C}(\omega) \neq 0.3$ for some $\omega \in \Omega$. If $A = B$, then:

$$\begin{aligned} G_D(A \cap B, B - A, A - B) &= 0 \text{ and} \\ G_D(C \cap B, B - C, C - B) &= G_D(C \cap A, A - C, C - A) = 1. \end{aligned}$$

The same happens when $B = C$. Otherwise, if $A \neq B$ and $B \neq C$, then both $G_D(C \cap B, B - C, C - B)$ and $G_D(A \cap B, B - A, A - B)$ are greater or equal to 0.5, and its sum equals 1.

Therefore, G_D generates a distance for IF-sets. To show that it generates neither an IF-dissimilarity nor an IF-divergence, it is enough to consider the IF-difference of Example 2.56, because in that case the function G_D generates the distance of Examples 5.10, where we showed that such function is neither an IF-dissimilarity nor an IF-divergence. \blacklozenge

We have seen sufficient conditions for G_D to generate distances, IF-dissimilarities and IF-divergences. However, such conditions are not necessary, and we cannot assure that every distance, IF-dissimilarity or IF-divergence can be generated in this way.

As we have seen, IF-divergences are more restrictive than IF-dissimilarities and IF-dissimilitudes. Thus, IF-divergences avoid some counterintuitive measures of comparison of IF-sets, since the stronger the conditions, the more “robust” the measure is. Because of this, we think it is preferable to work with IF-divergences, and we shall focus on them in the remainder of this chapter.

5.1.2 Properties of the IF-divergences

We have proposed an axiomatic definition of divergence measures for intuitionistic fuzzy sets, which are particular cases of dissimilarity and dissimilitude measures. Next, we study their properties in more detail. We begin by noting that a desirable property for a measure of the difference between IF-sets is positivity. Although it has not been imposed in the definition, it can be easily derived from axioms IF-Diss.1 and IF-Div.3:

Lemma 5.20 *If $D : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$ satisfies IF-Diss.1 and IF-Div.3, then it is positive.*

Proof: Consider two IF-sets A and B . From IF-Div.3, for every $C \in IFSs(\Omega)$ it holds that:

$$D(A, B) \geq D(A \cap C, B \cap C).$$

If we take $C = \emptyset$, then:

$$D(A, B) \geq D(A \cap \emptyset, B \cap \emptyset) = D(\emptyset, \emptyset) = 0,$$

by IF-Diss.1. Thus, D is a positive function. \blacksquare

Now we investigate an interesting property of IF-divergences.

Proposition 5.21 *Given an IF-divergence D_{IFS} , it fulfills that:*

$$D_{IFS}(A \cap B, B) = D_{IFS}(A, A \cup B),$$

and this value is lower than or equal to $D_{\text{IFS}}(A, B)$ and $D_{\text{IFS}}(A \cap B, A \cup B)$, that is:

$$D_{\text{IFS}}(A \cap B, B) = D_{\text{IFS}}(A, A \cup B) \leq \min\{D_{\text{IFS}}(A, B), D_{\text{IFS}}(A \cap B, A \cup B)\}.$$

However, there is no fixed relationship between $D_{\text{IFS}}(A \cap B, A \cup B)$ and $D_{\text{IFS}}(A, B)$.

Proof: By the definitions of union and intersection of intuitionistic fuzzy sets, we have that $(A \cup B) \cap B = B$ and $(A \cap B) \cup A = A$. Applying axioms IF-Div.3 and IF-Div.4, we obtain that

$$\begin{aligned} D_{\text{IFS}}(A \cap B, B) &= D_{\text{IFS}}(A \cap B, (A \cup B) \cap B) \leq D_{\text{IFS}}(A, A \cup B) \\ &= D_{\text{IFS}}((A \cap B) \cup A, B \cup A) \leq D_{\text{IFS}}(A \cap B, B). \end{aligned}$$

Thus, $D_{\text{IFS}}(A \cap B, B) = D_{\text{IFS}}(A, A \cup B)$.

On the other hand, $B \cap B = B$, whence

$$D_{\text{IFS}}(A \cap B, B) = D_{\text{IFS}}(A \cap B, B \cap B) \leq D_{\text{IFS}}(A, B) \text{ by Axiom IF-Div.3.}$$

Finally, since $A \cap B \subseteq A \subseteq A \cup B$, by Lemma 5.5 we have that

$$D_{\text{IFS}}(A, A \cup B) \leq D_{\text{IFS}}(A \cap B, A \cup B).$$

In order to prove that there is no dominance relationship between $D_{\text{IFS}}(A \cap B, A \cup B)$ and $D_{\text{IFS}}(A, B)$, let us consider the universe $\Omega = \{\omega\}$ and the IF-sets:

$$\left. \begin{array}{l} A = \{(\omega, 0.2, 0.6)\} \\ B = \{(\omega, 0.3, 0.7)\} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} A \cap B = \{(\omega, 0.2, 0.7)\} \\ A \cup B = \{(\omega, 0.3, 0.6)\} \end{array} \right.$$

Consider the IF-divergences D_L and l_{IFS} defined by:

$$\begin{aligned} D_L(A, B) &= \frac{1}{4} (|(\mu_A(\omega) - \nu_A(\omega)) - (\mu_B(\omega) - \nu_B(\omega))| + |\mu_A(\omega) - \mu_B(\omega)| \\ &\quad + |\nu_A(\omega) - \nu_B(\omega)|). \\ l_{\text{IFS}}(A, B) &= \frac{1}{2} |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + |\pi_A(\omega) - \pi_B(\omega)|. \end{aligned}$$

As we shall see in Section 5.1.3, they correspond to the Hong and Kim IF-divergence and the Hamming distance, respectively. Then:

$$\begin{aligned} l_{\text{IFS}}(A, B) &= 0.2 \text{ and} \\ l_{\text{IFS}}(A \cap B, A \cup B) &= 0.1. \\ D_L(A, B) &= \frac{0.2}{4} \text{ and} \\ D_L(A \cap B, A \cup B) &= \frac{0.2+0.1+0.1}{4} = \frac{0.4}{4}. \end{aligned}$$

Thus:

$$l_{\text{IFS}}(A, B) > l_{\text{IFS}}(A \cap B, A \cup B) \text{ and } D_L(A, B) < D_L(A \cap B, A \cup B)$$

and therefore, there is not fixed relationship between these two quantities. ■

Next, we shall study under which conditions axioms IF-Div.3 and IF-Div.4 are equivalent. But before tackling this problem, we give an example showing that they are not equivalent in general.

Example 5.22 Consider the function $D : IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}$ given by

$$D(A, B) = \sum_{\omega \in \Omega} h(\mu_A(\omega), \mu_B(\omega)), \text{ for every } A, B \in IFSS(\Omega),$$

where h is defined by

$$h(x, y) = \begin{cases} 0 & \text{if } x = y. \\ 1 - xy & \text{if } x \neq y. \end{cases}$$

We shall prove in Example 5.53 of Section 5.1.5 that D satisfies IF-Diss.1, IF-Diss.2 and IF-Div.4. However, it is not an IF-divergence. For instance, if we consider a universe $\Omega = \{\omega_1, \dots, \omega_n\}$, and the IF-sets defined by:

$$\begin{aligned} A &= \{(\omega, 0.2, 0.8) \mid \omega \in \Omega\}. \\ B &= \{(\omega, 0.8, 0.2) \mid \omega \in \Omega\}. \\ C &= \{(\omega, 0.5, 0.5) \mid \omega \in \Omega\} \end{aligned}$$

it holds that:

$$\begin{aligned} D(A \cap C, B \cap C) &= D(A, C) = \sum_{\omega \in \Omega} (1 - 0.2 \cdot 0.5) = \sum_{\omega \in \Omega} 0.9 = 0.9n. \\ D_{IF}(A, B) &= \sum_{\omega \in \Omega} (1 - 0.2 \cdot 0.8) = \sum_{\omega \in \Omega} 0.84 = 0.84n. \end{aligned}$$

Thus, $D(A \cap C, B \cap C) = 0.9n > 0.84n = D(A, B)$, and therefore IF-Div.3 is not satisfied.

Hence, we have an example of a function that satisfies IF-Div.4 but it does not satisfy IF-Div.3. Next we are going to show by means of an example that IF-Div.3 does not imply IF-Div.4 either. Consider the function $D : IFSS(\Omega) \times IFSS(\Omega) \rightarrow \mathbb{R}$ given by:

$$D(A, B) = \sum_{\omega \in \Omega} h(\mu_A(\omega), \mu_B(\omega)) \text{ for every } A, B \in FS(\Omega),$$

where $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ is defined by:

$$h(x, y) = \begin{cases} 0 & \text{if } x = y. \\ xy & \text{if } x \neq y. \end{cases}$$

We shall also see in Example 5.53 of Section 5.1.5 that this function satisfies IF-Diss.1, IF-Diss.2 and IF-Div.3, but it is not an IF-divergence: consider $\Omega = \{\omega_1, \dots, \omega_n\}$, and the IF-sets of the previous example. Then, it holds that

$$\begin{aligned} D(A \cup C, B \cup C) &= D(C, B) = \sum_{\omega \in \Omega} 0.8 \cdot 0.5 = \sum_{\omega \in \Omega} 0.4 = 0.4n. \\ D(A, B) &= \sum_{\omega \in \Omega} 0.2 \cdot 0.8 = \sum_{\omega \in \Omega} 0.16 = 0.16n. \end{aligned}$$

We can conclude that axiom IF-Div.4 is not satisfied since

$$D(A \cup C, B \cup C) = 0.4n > 0.16n = D(A, B). \blacklozenge$$

Therefore, axioms IF-Div.3 and IF-Div.4 are not related in general. We shall see however, that under some additional conditions they become equivalent. Let us consider the following natural property:

IF-Div.5: $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c)$ for every $A, B \in \text{IFSS}(\Omega)$.

In the following section we shall see some examples of IF-divergences satisfying this property. To see, however, that not all IF-divergences satisfy IF-Div.5, take $\Omega = \{\omega\}$ and the function defined by:

$$D_{\text{IFS}}(A, B) = |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|^2. \quad (5.2)$$

We shall prove in Example 5.54 of Section 5.1.5 that this function is an IF-divergence. However, it does not satisfy IF-Div.5. To see that, consider the IF-sets

$$A = \{(\omega, 0.6, 0.4)\} \text{ and } B = \{(\omega, 0.5, 0.1)\}.$$

It holds that:

$$D_{\text{IFS}}(A, B) = 0.1 + 0.09 = 0.19 \neq 0.31 = 0.3 + 0.01 = D_{\text{IFS}}(A^c, B^c).$$

Our next result shows that, when IF-Div.5 is satisfied, then axioms IF-Div.3 and IF-Div.4 are equivalent.

Proposition 5.23 *If D is a function $D : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ satisfying the property IF-Div.5, then it satisfies IF-Div.3 if and only if it satisfies IF-Div.4.*

Proof: First of all let us show that, since $D(A, B) = D(A^c, B^c)$ by IF-Div.5, it also holds that:

$$D(A \cup C, B \cup C) = D((A \cup C)^c, (B \cup C)^c) = D(A^c \cap C^c, B^c \cap C^c).$$

Assume that D satisfies IF-Div.3:

$$D(A \cap C, B \cap C) \leq D(A, B) \text{ for every } A, B \in \text{IFSS}(\Omega).$$

Then it also satisfies IF-Div.4:

$$D(A \cup C, B \cup C) = D(A^c \cap C^c, B^c \cap C^c) \leq D(A^c, B^c) = D(A, B).$$

Similarly, assume that D satisfies IF-Div.4, that is,

$$D(A \cup C, B \cup C) \leq D(A, B) \text{ for every } A, B \in \text{IFSS}(\Omega).$$

Then, it also satisfies axiom IF-Div.3:

$$D(A \cap C, B \cap C) = D(A^c \cup C^c, B^c \cup C^c) \leq D(A^c, B^c) = D(A, B). \quad \blacksquare$$

Now, we will obtain a general expression of IF-divergences by comparing the membership and non-membership functions of the IF-sets by means of a t-conorm.

Proposition 5.24 Consider a finite set Ω . If S and S^* are two t -conorms, the function D_{IFS} defined by:

$$D_{\text{IFS}}(A, B) = S_{\omega \in \Omega}(S^*(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|))$$

for every $A, B \in \text{IFSS}(\Omega)$, is an IF-divergence. Moreover, it satisfies IF-Div.5.

Proof: Let us prove that D_{IFS} fulfills axioms IF-Diss.1 to IF-Div.4.

IF-Diss.1: Let A be an IF-set. Obviously, $D_{\text{IFS}}(A, A) = 0$:

$$D_{\text{IFS}}(A, A) = S_{\omega \in \Omega}(S^*(0, 0)) = S(0, \dots, 0) = 0.$$

IF-Diss.2: Let A and B be two IF-sets. It holds that:

$$\begin{aligned} D_{\text{IFS}}(A, B) &= S_{\omega \in \Omega}(S^*(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|)) \\ &= S_{\omega \in \Omega}(S^*(|\mu_B(\omega) - \mu_A(\omega)|, |\nu_B(\omega) - \nu_A(\omega)|)) = D_{\text{IFS}}(B, A). \end{aligned}$$

IF-Div.3: Let A, B and C three IF-sets. We have to prove that

$$D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cap C, B \cap C).$$

Applying the first part of Lemma A.1 of Appendix A, we have that

$$\begin{aligned} |\mu_A(\omega) - \mu_B(\omega)| &\geq |\min(\mu_A(\omega), \mu_C(\omega)) - \min(\mu_B(\omega), \mu_C(\omega))| = |\mu_{A \cap C}(\omega) - \mu_{B \cap C}(\omega)|. \\ |\nu_A(\omega) - \nu_B(\omega)| &\geq |\max(\nu_A(\omega), \nu_C(\omega)) - \max(\nu_B(\omega), \nu_C(\omega))| = |\nu_{A \cap C}(\omega) - \nu_{B \cap C}(\omega)|. \end{aligned}$$

Since every t -conorm is increasing, it holds that:

$$\begin{aligned} D_{\text{IFS}}(A, B) &= S_{\omega \in \Omega}(S^*(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|)) \\ &\geq S_{\omega \in \Omega}(S^*(|\mu_{A \cap C}(\omega) - \mu_{B \cap C}(\omega)|, |\nu_{A \cap C}(\omega) - \nu_{B \cap C}(\omega)|)) \\ &= D_{\text{IFS}}(A \cap C, B \cap C). \end{aligned}$$

IF-Div.4: Consider three IF-sets A, B and C . Using the first part of Lemma A.1 of Appendix A, we see that:

$$\begin{aligned} |\mu_A(\omega) - \mu_B(\omega)| &\geq |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))| \\ &= |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|. \\ |\nu_A(\omega) - \nu_B(\omega)| &\geq |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))| \\ &= |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|. \end{aligned}$$

Since t -conorms are increasing operators,

$$\begin{aligned} D_{\text{IFS}}(A, B) &= S_{\omega \in \Omega}(S^*(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|)) \\ &\geq S_{\omega \in \Omega}(S^*(|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|, |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|)) \\ &= D_{\text{IFS}}(A \cup C, B \cup C). \end{aligned}$$

Thus, D_{IFS} is an IF-divergence. Now, we are going to prove that it also satisfies IF-Div.5. Using that every t-conorm is symmetric, we deduce that:

$$\begin{aligned} D_{\text{IFS}}(A, B) &= S_{\omega \in \Omega}(S^*(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|)) \\ &= S_{\omega \in \Omega}(S^*(|\nu_A(\omega) - \nu_B(\omega)|, |\mu_A(\omega) - \mu_B(\omega)|)) = D_{\text{IFS}}(A^c, B^c). \end{aligned}$$

Therefore $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c)$ for every $A, B \in \text{IFSS}(\Omega)$. \blacksquare

One of the conditions we required on IF-divergences was that “the more similar two IF-sets are, the lower the divergence is between them”. In the following result we are going to see that, if the non-membership functions of A and B are the same than the ones of C and D , respectively, or the membership functions of C and D are the same, then the IF-divergence between A and B is greater than the IF-divergence between C and D .

Proposition 5.25 *Let A and B be two IF-sets. Let us consider the IF-sets C_A and D_B given by:*

$$\begin{aligned} C_A &= \{(\omega, \mu(\omega), \nu_A(\omega)) \mid \omega \in \Omega\}, \\ D_B &= \{(\omega, \mu(\omega), \nu_B(\omega)) \mid \omega \in \Omega\}, \end{aligned}$$

where $\mu : \Omega \rightarrow [0, 1]$ is a map such that $\mu(\omega) + \nu_A(\omega) \leq 1$ and $\mu(\omega) + \nu_B(\omega) \leq 1$ for every $\omega \in \Omega$. If D is an IF-divergence, then $D(A, B) \geq D(C_A, D_B)$.

Proof: Let us define the following IF-set:

$$N = \{(\omega, \min(\mu_A(\omega), \mu(\omega)), 0) \mid \omega \in \Omega\}.$$

Then,

$$\begin{aligned} A \cap N &= \{(\omega, \min(\mu_A(\omega), \mu(\omega)), \nu_A(\omega)) \mid \omega \in \Omega\}. \\ B \cap N &= \{(\omega, \min(\mu_B(\omega), \mu(\omega), \mu_A(\omega)), \nu_B(\omega)) \mid \omega \in \Omega\}. \end{aligned}$$

Applying IF-Div.3 we obtain that $D(A, B) \geq D(A \cap N, B \cap N)$. Consider now another IF-set, defined by:

$$M = \{(\omega, \mu(\omega), \max(\nu_A(\omega), \nu_B(\omega))) \mid \omega \in \Omega\}.$$

We obtain that:

$$\begin{aligned} (A \cap N) \cup M &= \{(\omega, \max(\mu(\omega), \min(\mu_A(\omega), \mu(\omega))), \nu_A(\omega)) \mid \omega \in \Omega\} \\ &= \{(\omega, \mu(\omega), \nu_A(\omega)) \mid \omega \in \Omega\} = C_A. \\ (B \cap N) \cup M &= \{(\omega, \max(\mu(\omega), \min(\mu_B(\omega), \mu(\omega), \mu_A(\omega))), \nu_B(\omega)) \mid \omega \in \Omega\} \\ &= \{(\omega, \mu(\omega), \nu_B(\omega)) \mid \omega \in \Omega\} = D_B. \end{aligned}$$

Applying IF-Div.4,

$$D(A, B) \geq D(A \cap N, B \cap N) \geq D((A \cap N) \cup M, (B \cap N) \cup M) = D(C_A, D_B). \quad \blacksquare$$

Analogously, we can obtain a similar result by exchanging the membership and the non-membership functions.

Proposition 5.26 *Let A and B be two IF-sets. Let us consider the IF-sets C_A and D_B given by:*

$$C = \{(\omega, \mu_A(\omega), \nu(\omega)) \mid \omega \in \Omega\} \text{ and } D = \{(\omega, \mu_B(\omega), \nu(\omega)) \mid \omega \in \Omega\},$$

where $\nu : \Omega \rightarrow [0, 1]$ is a map such that $\mu_A(\omega) + \nu(\omega) \leq 1$ and $\mu_B(\omega) + \nu(\omega) \leq 1$ for every $\omega \in \Omega$. If D_{IFS} is an IF-divergence, then $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(C_A, D_B)$.

We conclude this section with a property that assures that some transformations of IF-divergences are also IF-divergences.

Proposition 5.27 *If D is an IF-divergence and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is an increasing function with $\phi(0) = 0$, then D^ϕ defined by:*

$$D_{\text{IFS}}^\phi(A, B) = \phi(D_{\text{IFS}}(A, B)) \text{ for every } A, B \in \text{IFSs}(\Omega),$$

is also an IF-divergence. Moreover, if D_{IFS} satisfies axiom IF-Div.5, then so does D_{IFS}^ϕ .

Proof: Let D_{IFS} be an IF-divergence and ϕ an increasing function with $\phi(0) = 0$. Condition IF-Diss.1 follows from $\phi(0) = 0$ and conditions IF-Div.3 and IF-Div.4 follow from the monotonicity of ϕ , and IF-Div.2 and IF-Div.5 are trivially fulfilled by definition. ■

5.1.3 Examples of IF-divergences and IF-dissimilarities

This subsection is devoted to the study of some of the most important examples of IF-divergences and dissimilarities. Specifically, we shall investigate whether the most prominent examples of dissimilarities that can be found in the literature are particular cases of IF-divergence. Furthermore, we shall also study if they satisfy other properties, such as axiom IF-Div.5, or if they are dissimilitudes.

Dissimilarities that also are IF-divergences

In this section we are going to present an overview of the dissimilarities that are also IF-divergences. From now on, Ω denotes a finite universe with n elements.

Hamming and normalized Hamming distance One of the most important comparison measures for IF-sets are the Hamming distance ([193]), defined by:

$$l_{\text{IFS}}(A, B) = \frac{1}{2} \sum_{\omega \in \Omega} (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + |\pi_A(\omega) - \pi_B(\omega)|),$$

and the normalized Hamming distance by:

$$l_{\text{nIFS}}(A, B) = \frac{1}{n} l_{\text{IFS}}(A, B), \text{ for every } A, B \in \text{IFSs}(\Omega).$$

These functions are known to be dissimilarities. Let us prove that they are also IF-divergences. In order to do this, we shall first of all prove that the Hamming distance is an IF-divergence; this, together with Proposition 5.27, will allow us to conclude that the normalized Hamming distance is also an IF-divergence, because it is an increasing transformation (by means of $\phi(x) = \frac{x}{n}$) of the Hamming distance. In order to prove that the Hamming distance is an IF-divergence, we shall begin by showing that it satisfies axiom IF-Div.5. Let us note that

$$\pi_A(\omega) = 1 - \mu_A(\omega) - \nu_A(\omega) = 1 - \nu_{A^c}(\omega) - \mu_{A^c}(\omega) = \pi_{A^c}(\omega)$$

for every $\omega \in \Omega$ and $A \in \text{IFSs}(\Omega)$. Then:

$$\begin{aligned} l_{\text{IFS}}(A^c, B^c) &= \sum_{\omega \in \Omega} (|\nu_A(\omega) - \nu_B(\omega)| + |\mu_A(\omega) - \mu_B(\omega)| + |\pi_{A^c}(\omega) - \pi_{B^c}(\omega)|) \\ &= \sum_{\omega \in \Omega} (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + |\pi_A(\omega) - \pi_B(\omega)|) = l_{\text{IFS}}(A, B). \end{aligned}$$

By Proposition 5.23, axioms IF-Div.3 and IF-Div.4 are equivalent. Moreover, axioms IF-Diss.1 and IF-Diss.2 are satisfied since l_{IFS} is an IF-dissimilarity (see for instance [92]). Hence, in order to prove that l_{IFS} is an IF-divergence it suffices to check that it fulfills either IF-Div.3 or IF-Div.4. Let us show the latter. Let A , B and C be three IF-sets; using Lemma A.2 of Appendix A, we know that for every $\omega \in \Omega$, the following inequality holds:

$$\begin{aligned} &|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + |\pi_A(\omega) - \pi_B(\omega)| \geq \\ &|\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))| + \\ &|\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))| + \\ &|\max(\mu_A(\omega), \mu_C(\omega)) + \min(\nu_A(\omega), \nu_C(\omega)) - \\ &\max(\mu_B(\omega), \mu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))|. \end{aligned}$$

Then:

$$\begin{aligned} l_{\text{IFS}}(A, B) &= \sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + |\pi_A(\omega) - \pi_B(\omega)| \\ &\geq \sum_{\omega \in \Omega} |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))| \\ &\quad + |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))| \\ &\quad + |\max(\mu_A(\omega), \mu_C(\omega)) - \min(\nu_A(\omega), \nu_C(\omega)) \\ &\quad + \max(\mu_B(\omega), \mu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))| = l_{\text{IFS}}(A \cup C, B \cup C). \end{aligned}$$

Thus, $l_{\text{IFS}}(A, B) \geq l_{\text{IFS}}(A \cup C, B \cup C)$.

In other words, we have proven that l_{IFS} satisfies axiom IF-Div.4, and therefore it also satisfies IF-Div.3. Hence, l_{IFS} is an IF-divergence, and as a consequence so is l_{nIFS} .

Moreover, since they are IF-divergences, we deduce that they are also dissimilitudes. In summary, the Hamming and the normalized Hamming distances are examples of dissimilarities, IF-divergences, dissimilitudes and distances.

Hausdorff dissimilarity Another very important dissimilarity between IF-sets is based on the Hausdorff distance (see for example [85]). It is defined by:

$$d_H(A, B) = \sum_{\omega \in \Omega} \max(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|).$$

As the Hamming distance, the Hausdorff dissimilarity satisfies axiom IF-Div.5, because

$$d_H(A^c, B^c) = \sum_{\omega \in \Omega} \max(|\nu_A(\omega) - \nu_B(\omega)|, |\mu_A(\omega) - \mu_B(\omega)|) = d_H(A, B).$$

Applying Prop 5.23, we deduce that axioms IF-Div.3 and IF-Div.4 are equivalent. Note that axioms IF-Diss.1 and IF-Diss.2 are satisfied by d_H since it is a IF-dissimilarity. Hence, in order to prove that d_H is an IF-divergence, it suffices to prove that either IF-Div.3 or IF-Div.4 hold.

Let us prove that axiom IF-Div.4 is satisfied by d_H . Consider three IF-sets A, B and C . Then, the IF-sets $A \cup C$ and $B \cup C$ are given by:

$$\begin{aligned} A \cup C &= \{(\omega, \max(\mu_A(\omega), \mu_C(\omega)), \min(\nu_A(\omega), \nu_C(\omega))) \mid \omega \in \Omega\}. \\ B \cup C &= \{(\omega, \max(\mu_B(\omega), \mu_C(\omega)), \min(\nu_B(\omega), \nu_C(\omega))) \mid \omega \in \Omega\}. \end{aligned}$$

By the second part of Lemma A.1 of Appendix A, it holds that:

$$\begin{aligned} |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))| &\leq |\mu_A(\omega) - \mu_B(\omega)|. \\ |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))| &\leq |\nu_A(\omega) - \nu_B(\omega)|. \end{aligned}$$

Then,

$$\begin{aligned} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| &\leq |\mu_A(\omega) - \mu_B(\omega)| \text{ and} \\ |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)| &\leq |\nu_A(\omega) - \nu_B(\omega)|. \end{aligned}$$

From these inequalities it follows that:

$$\begin{aligned} \max(|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|, |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|) \\ \leq \max(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|). \end{aligned}$$

This inequality has been proved for every ω in Ω , and consequently:

$$\begin{aligned} d_H(A \cup C, B \cup C) &= \sum_{\omega \in \Omega} \max(|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|, |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|) \\ &\leq \sum_{\omega \in \Omega} \max(|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|) = d_H(A, B). \end{aligned}$$

Thus, the Hausdorff IF-dissimilarity is an IF-divergence, and consequently it is also a dissimilitude.

Note that it is also possible to define the normalized Hausdorff dissimilarity, denoted by d_{nH} , by:

$$d_{nH}(A, B) = \frac{1}{n} d_H(A, B), \text{ for every } A, B \in IFSs(\Omega).$$

It holds that $d_{nH}(A, B) = \phi(d_H(A, B))$, where $\phi(x) = \frac{1}{n}x$. As we already said, this function ϕ is increasing and $\phi(0) = 0$. Therefore, using Proposition 5.27, we deduce that d_{nH} is also an IF-divergence that fulfills axiom IF-Div.5.

We conclude that d_H and d_{nH} are distances, IF-dissimilarities, IF-divergences and IF-dissimilitudes at the same time.

Hong & Kim dissimilarities Hong and Kim proposed two dissimilarity measures in [89]. They are defined by:

$$D_C(A, B) = \frac{1}{2n} \sum (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|) \text{ and}$$

$$D_L(A, B) = \frac{1}{4n} \left(\sum_{\omega \in \Omega} |S_A(\omega) - S_B(\omega)| + \left(\sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| \right) \right),$$

where $S_A(\omega) = \mu_A(\omega) - \nu_A(\omega)$ and $S_B(\omega) = \mu_B(\omega) - \nu_B(\omega)$.

Recall that D_L can be equivalently expressed by:

$$D_L(A, B) = \frac{1}{4n} \sum_{\omega \in \Omega} |(\mu_A(\omega) - \mu_B(\omega)) - (\nu_A(\omega) - \nu_B(\omega))| + |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|$$

for every $A, B \in IFSs(\Omega)$.

In order to prove that D_C satisfies IF-Div.3, we shall use part b) of Lemma A.1:

$$|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| \geq$$

$$|\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))| + |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))|.$$

Using this fact, IF-Div.3 trivially follows, and IF-Div.4 can be similarly proved.

Let us see that D_L is also an IF-divergence. For this, it suffices to take into account that, from Lemma A.3, for every $\omega \in \Omega$ it holds that:

$$|\mu_A(\omega) - \mu_B(\omega) - \nu_A(\omega) + \nu_B(\omega)| + |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|$$

$$\geq |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))$$

$$- \min(\nu_A(\omega), \nu_C(\omega)) + \min(\nu_B(\omega), \nu_C(\omega))|$$

$$+ |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))|$$

$$+ |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))|.$$

By taking the sum on Ω on every part of the inequality, and multiplying each term by $\frac{1}{4n}$, we obtain that:

$$D_L(A, B) \geq D_L(A \cup C, B \cup C).$$

Thus, D_L satisfies axiom IF-Div.4, and therefore also IF-Div.3 since D_L satisfies the property IF-Div.5. We conclude that both D_C and D_L are IF-dissimilarities, IF-divergences and IF-dissimilitudes.

Li et al. dissimilarity Another dissimilarity measure for IF-sets was proposed by Li et al. ([113]):

$$D_O(A, B) = \frac{1}{\sqrt{2n}} \left(\sum_{\omega \in \Omega} (\mu_A(\omega) - \mu_B(\omega))^2 + (\nu_A(\omega) - \nu_B(\omega))^2 \right)^{\frac{1}{2}}.$$

This dissimilarity also satisfies IF-Div.5, since $D_O(A^c, B^c) = D_O(A, B)$. Then, by Proposition 5.23, in order to prove that D_O is an IF-divergence it is enough to prove that it satisfies IF-Div.4. Let us consider A , B and C three IF-sets. By the second part of Lemma A.1 in Appendix A, we know that:

$$\begin{aligned} |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))| &\leq |\mu_A(\omega) - \mu_B(\omega)| \text{ and} \\ |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))| &\leq |\nu_A(\omega) - \nu_B(\omega)|, \end{aligned}$$

or, equivalently,

$$\begin{aligned} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| &\leq |\mu_A(\omega) - \mu_B(\omega)| \text{ and} \\ |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)| &\leq |\nu_A(\omega) - \nu_B(\omega)|. \end{aligned}$$

Then it holds that:

$$\begin{aligned} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|^2 + |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|^2 \\ \leq |\mu_A(\omega) - \mu_B(\omega)|^2 + |\nu_A(\omega) - \nu_B(\omega)|^2, \end{aligned}$$

whence

$$\begin{aligned} D_O(A \cup C, B \cup C) &= \frac{1}{\sqrt{2n}} \left(\sum_{\omega \in \Omega} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|^2 + |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|^2 \right)^{\frac{1}{2}} \\ &\leq \frac{1}{\sqrt{2n}} \left(\sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)|^2 + |\nu_A(\omega) - \nu_B(\omega)|^2 \right)^{\frac{1}{2}} = D_O(A, B). \end{aligned}$$

Thus, D_O satisfies axiom IF-Div.4 and therefore it is an IF-Divergence, and in particular an IF-dissimilitude.

Mitchell dissimilarity Mitchell ([138]) proposed a dissimilarity defined by:

$$D_{\text{HB}}(A, B) = \frac{1}{2\sqrt[p]{n}} \left(\left(\sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)|^p \right)^{\frac{1}{p}} + \left(\sum_{\omega \in \Omega} |\nu_A(\omega) - \nu_B(\omega)|^p \right)^{\frac{1}{p}} \right),$$

for some $p \geq 1$. This dissimilarity obviously satisfies IF-Div.5. Thus, in order to prove that D_{HB} is an IF-divergence it is enough to prove IF-Div.4, since IF-Diss.1 and IF-Diss.2 are satisfied for every dissimilarity. Consider A , B and C . Applying again the second part of Lemma A.1 from Appendix A we deduce that:

$$\begin{aligned} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| &\leq |\mu_A(\omega) - \mu_B(\omega)| \text{ and} \\ |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)| &\leq |\nu_A(\omega) - \nu_B(\omega)|. \end{aligned}$$

Moreover, the inequalities holds if we raise every term to the power of p , whence

$$\begin{aligned} D_{\text{HB}}(A, B) &= \frac{1}{2\sqrt[p]{n}} \left(\left(\sum_{\omega \in \Omega} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)|^p \right)^{\frac{1}{p}} + \left(\sum_{\omega \in \Omega} |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|^p \right)^{\frac{1}{p}} \right) \\ &\leq \frac{1}{2\sqrt[p]{n}} \left(\left(\sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)|^p \right)^{\frac{1}{p}} + \left(\sum_{\omega \in \Omega} |\nu_A(\omega) - \nu_B(\omega)|^p \right)^{\frac{1}{p}} \right) \\ &= D_{\text{HB}}(A, B). \end{aligned}$$

Thus, axiom IF-Div.4 holds, and therefore D_{HB} is an IF-divergence, and in particular a dissimilitude.

Liang & Shi dissimilarities Liang and Shi ([114]) defined the dissimilarities D_e^p and D_h^p , for some $p \geq 1$, by

$$\begin{aligned} D_e^p(A, B) &= \frac{1}{2\sqrt[p]{n}} \left(\sum_{\omega \in \Omega} (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|)^p \right)^{\frac{1}{p}}, \\ D_h^p(A, B) &= \frac{1}{\sqrt[p]{3n}} \left(\sum_{\omega \in \Omega} (\eta_1(\omega) + \eta_2(\omega) + \eta_3(\omega))^p \right)^{\frac{1}{p}}, \end{aligned}$$

where

$$\begin{aligned} \eta_1(\omega) &= \frac{1}{2} (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|), \\ \eta_2(\omega) &= \frac{1}{2} |\mu_A(\omega) - \nu_A(\omega) - \mu_B(\omega) + \nu_B(\omega)|, \\ \eta_3(\omega) &= \max(l_A(\omega), l_B(\omega)) - \min(l_A(\omega), l_B(\omega)), \\ l_A(\omega) &= \frac{1}{2} (1 - \nu_A(\omega) - \mu_A(\omega)), \\ l_B(\omega) &= \frac{1}{2} (1 - \nu_B(\omega) - \mu_B(\omega)). \end{aligned}$$

Note that D_h^p can be expressed in a equivalent way as

$$\begin{aligned} D_h^p(A, B) &= \frac{1}{2\sqrt[p]{3n}} \left(\sum_{\omega \in \Omega} (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| \right. \\ &\quad \left. + |(\mu_A(\omega) - \mu_B(\omega)) - (\nu_A(\omega) - \nu_B(\omega))| \right. \\ &\quad \left. + |(\mu_A(\omega) + \nu_A(\omega)) - (\mu_B(\omega) + \nu_B(\omega))|)^p \right)^{\frac{1}{p}}. \end{aligned}$$

As in the previous examples, both D_e^p and D_h^p satisfy IF-Div.5, and therefore it suffices to prove that both functions satisfy IF-Div.4 to prove that they are IF-divergences. Let us first focus on D_e^p , and let us consider A , B and C three IF-sets. Applying again the second part of Lemma A.1 in Appendix A we know that:

$$\begin{aligned} |\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| &\leq |\mu_A(\omega) - \mu_B(\omega)| \text{ and} \\ |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)| &\leq |\nu_A(\omega) - \nu_B(\omega)|. \end{aligned}$$

If we sum both inequalities we obtain

$$|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| + |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)| \leq |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|,$$

and since this inequality also holds when we raise every component to the power of p ,

$$\begin{aligned} D_e^p(A \cup C, B \cup C) &= \frac{1}{2^{\frac{1}{p}n}} \left(\sum_{\omega \in \Omega} (|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| + |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)|)^p \right)^{\frac{1}{p}} \\ &\leq \frac{1}{2^{\frac{1}{p}n}} \left(\sum_{\omega \in \Omega} (|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|)^p \right)^{\frac{1}{p}} = D_e^p(A, B). \end{aligned}$$

Thus, D_e^p satisfies IF-Div.4, and, taking into account that it satisfies IF-Div.5, also axiom IF-Div.3. Hence, it is a dissimilarity, and consequently, a dissimilitude.

Consider now D_h^p . Using Lemma A.4 in Appendix A, we know that, for every $\omega \in \Omega$,

$$\begin{aligned} &|\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + \\ &|\mu_A(\omega) - \mu_B(\omega) - \nu_A(\omega) + \nu_B(\omega)| + \\ &|\mu_A(\omega) + \nu_A(\omega) - \mu_B(\omega) - \nu_B(\omega)| \geq \\ &|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega)| + |\nu_{A \cup C}(\omega) - \nu_{B \cup C}(\omega)| + \\ &|\mu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega) - \nu_{A \cup C}(\omega) + \nu_{B \cup C}(\omega)| + \\ &|\mu_{A \cup C}(\omega) + \nu_{A \cup C}(\omega) - \mu_{B \cup C}(\omega) - \nu_{B \cup C}(\omega)|. \end{aligned}$$

Making the summation over every ω in Ω in each part of the inequality and multiplying by $\frac{1}{2^{\frac{1}{p}3n}}$, we obtain that $D_h^p(A, B) \geq D_h^p(A \cup C, B \cup C)$.

Thus, both D_e^p and D_h^p are IF-dissimilarities, IF-divergences and IF-dissimilitudes.

Hung & Yang dissimilarities Hung and Yang proposed some new dissimilarities in [92], two of which are based on the Hausdorff dissimilarity. As we shall see, it is easy to check that both are also IF-divergences. These dissimilarities are defined by:

$$\begin{aligned} D_{HY}^1(A, B) &= d_{nH}(A, B). \\ D_{HY}^2(A, B) &= 1 - \frac{e^{-d_{nH}(A, B)} - e^{-1}}{1 - e^{-1}}. \\ D_{HY}^3(A, B) &= 1 - \frac{1 - d_{nH}(A, B)}{1 + d_{nH}(A, B)}. \end{aligned}$$

We have already proven that the Hausdorff dissimilarity is an IF-divergence that satisfies the property IF-Div.5. Consider the functions ϕ_2 and ϕ_3 defined by:

$$\phi_2(x) = 1 - \frac{e^{-x} - e^{-1}}{1 - e^{-1}} \text{ and } \phi_3(x) = 1 - \frac{1-x}{1+x}.$$

These functions are increasing and satisfy $\phi_2(0) = \phi_3(0) = 0$. Applying Proposition 5.27 we conclude that

$$\begin{aligned} d_H^{\phi_2}(A, B) &= \phi_2(d_{\text{nH}}(A, B)) = D_{\text{HY}}^2(A, B) \text{ and} \\ d_H^{\phi_3}(A, B) &= \phi_3(d_{\text{nH}}(A, B)) = D_{\text{HY}}^3(A, B) \end{aligned}$$

are IF-divergences that satisfy property IF-Div.5. Thus, they are also IF-dissimilitudes.

On the other hand, Hung and Yang also proposed the IF-dissimilarity given by

$$D_{\text{pk2}}(A, B) = \frac{1}{2} \left(\max_{\omega \in \Omega} (|\mu_A(\omega) - \mu_B(\omega)|) + \max_{\omega \in \Omega} (|\nu_A(\omega) - \nu_B(\omega)|) \right).$$

This measure satisfies IF-Div.5, whence, applying Proposition 5.23, it is enough to prove that, indeed, D_{pk2} satisfies IF-Div.4. If we consider A , B and C three IF-sets, we know from the second part of Lemma A.1 in Appendix A that:

$$\begin{aligned} |\mu_A(\omega) - \mu_B(\omega)| &\geq |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))|. \\ |\nu_A(\omega) - \nu_B(\omega)| &\geq |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))|. \end{aligned}$$

Thus,

$$\begin{aligned} \max_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)| &\geq \max_{\omega \in \Omega} |\max(\mu_A(\omega), \mu_C(\omega)) - \max(\mu_B(\omega), \mu_C(\omega))|. \\ \max_{\omega \in \Omega} |\nu_A(\omega) - \nu_B(\omega)| &\geq \max_{\omega \in \Omega} |\min(\nu_A(\omega), \nu_C(\omega)) - \min(\nu_B(\omega), \nu_C(\omega))|. \end{aligned}$$

Then, $D_{\text{pk2}}(A, B) \geq D_{\text{pk2}}(A \cup C, B \cup C)$. We conclude that D_{pk2} is another example of IF-dissimilarity that is also an IF-divergence and IF-dissimilitude.

Dissimilarities that are not IF-divergences

Let us now provide some examples of dissimilarities, very frequently used in the literature, that are not IF-divergences. We shall also give some examples showing that these comparison measures are, in some cases, counterintuitive.

Euclidean and normalized Euclidean distance Together with the Hamming and Hausdorff distances, one of the most important comparison measures is the Euclidean

distance (see for example. nThis distance is used to define a dissimilarity between IF-sets and its normalization as follows ([85]):

$$q_{\text{IFS}}(A, B) = \left(\frac{1}{2} \sum_{\omega \in \Omega} (\mu_A(\omega) - \mu_B(\omega))^2 + (\nu_A(\omega) - \nu_B(\omega))^2 + (\pi_A(\omega) - \pi_B(\omega))^2 \right)^{\frac{1}{2}}.$$

$$q_{\text{nIFS}}(A, B) = \frac{1}{n} q_{\text{IFS}}(A, B).$$

These dissimilarities fulfill axiom IF-Div.5, since $\pi_A(\omega) = \pi_{A^c}(\omega)$ and $\pi_B(\omega) = \pi_{B^c}(\omega)$ for every $A, B \in \text{IFSS}(\Omega)$. However, they are not IF-divergences, since they do not satisfy axioms IF-Div.3 nor IF-Div.4. To see a counterexample, consider $\Omega = \{\omega\}$ and the following IF-sets:

$$A = \{(\omega, 0.12, 0.68)\}, B = \{(\omega, 0.29, 0.59)\}, C = \{(\omega, 0.11, 0.36)\}.$$

The IF-sets $A \cup C$ and $B \cup C$ are given by:

$$A \cup C = \{(\omega, 0.12, 0.36)\} \text{ and } B \cup C = \{(\omega, 0.29, 0.36)\}.$$

It holds that $q_{\text{IFS}}(A \cup C, B \cup C) > q_{\text{IFS}}(A, B)$:

$$q_{\text{IFS}}(A \cup C, B \cup C) = \left[\frac{1}{2} (0.17^2 + 0 + 0.17^2) \right]^{0.5} = 0.17.$$

$$q_{\text{IFS}}(A, B) = \left[\frac{1}{2} (0.17^2 + 0.09^2 + 0.08^2) \right]^{0.5} = 0.1473.$$

Moreover, since q_{IFS} does not satisfy IF-Div.4, axiom IF-Div.3 cannot hold either because they are equivalent under IF-Div.5. Therefore, q_{IFS} is neither an IF-divergence nor a dissimilitude. The same example shows that q_{nIFS} is not an IF-divergence, since for $n = 1$ we have that $q_{\text{IFS}} = q_{\text{nIFS}}$.

Liang & Shi dissimilarity We have seen previously some IF-dissimilarities proposed by Liang and Shi that are also IF-divergences. They also proposed another IF-dissimilarity measure, that is defined by:

$$D_s^p(A, B) = \frac{1}{\sqrt[p]{n}} \left(\sum_{\omega \in \Omega} (\varphi_{s1}(\omega) + \varphi_{s2}(\omega))^p \right)^{\frac{1}{p}},$$

where $p \geq 1$ and

$$\begin{aligned} \varphi_{s1}(\omega) &= \frac{1}{2} |m_{A1}(\omega) - m_{B1}(\omega)|. \\ \varphi_{s2}(\omega) &= \frac{1}{2} |m_{A2}(\omega) - m_{B2}(\omega)|. \\ m_{A1}(\omega) &= \frac{1}{2} (\mu_A(\omega) + m_A(\omega)). \\ m_{A2}(\omega) &= \frac{1}{2} (m_A(\omega) + 1 - \nu_A(\omega)). \\ m_{B1}(\omega) &= \frac{1}{2} (\mu_B(\omega) + m_B(\omega)). \\ m_{B2}(\omega) &= \frac{1}{2} (m_B(\omega) + 1 - \nu_B(\omega)). \\ m_A(\omega) &= \frac{1}{2} (\mu_A(\omega) + 1 - \nu_A(\omega)). \\ m_B(\omega) &= \frac{1}{2} (\mu_B(\omega) + 1 - \nu_B(\omega)). \end{aligned}$$

Note that D_s^p can also be expressed by:

$$D_s^p(A, B) = \frac{1}{\sqrt[p]{n}} \left(\sum_{\omega \in \Omega} \frac{1}{8} (|3(\mu_A(\omega) - \mu_B(\omega)) - (\nu_A(\omega) - \nu_B(\omega))| + |(\mu_A(\omega) - \mu_B(\omega)) - 3(\nu_A(\omega) - \nu_B(\omega))|) \right)^{\frac{1}{p}}.$$

Thus, this dissimilarity satisfies axiom IF-Div.5. However, neither IF-Div.3 nor IF-Div.4 are satisfied. To see this, consider $\Omega = \{\omega\}$ and the IF-sets

$$A = \{(\omega, 0.25, 0.25)\} \text{ and } B = \{(\omega, 0.6, 0.35)\}.$$

For these IF-sets it holds that $D_s^p(A, B) = 0.125$. Furthermore, if we consider the IF-set C defined by:

$$C = \{(\omega, 0.2, 0.2)\},$$

it holds that

$$A \cup C = \{(\omega, 0.25, 0.2)\} \text{ and } B \cup C = \{(\omega, 0.6, 0.2)\},$$

whence,

$$D_s^p(A \cup C, B \cup C) = 0.175 > 0.125 = D(A, B).$$

Consequently, D_s^p is neither an IF-divergence, nor an IF-dissimilitude.

Chen dissimilarity Chen ([36, 37]) defined an IF-dissimilarity measure by:

$$D_C(A, B) = \frac{1}{2n} \sum_{\omega \in \Omega} |S_A(\omega) - S_B(\omega)|,$$

where $S_A(\omega) = \mu_A(\omega) - \nu_A(\omega)$ and $S_B(\omega) = \mu_B(\omega) - \nu_B(\omega)$.

This dissimilarity also satisfies axiom IF-Div.5, because:

$$\begin{aligned} D_C(A^c, B^c) &= \frac{1}{2n} \sum_{\omega \in \Omega} |S_{A^c}(\omega) - S_{B^c}(\omega)| \\ &= \frac{1}{2n} \sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega) - \nu_A(\omega) + \nu_B(\omega)| \\ &= \frac{1}{2n} \sum_{\omega \in \Omega} |S_A(\omega) - S_B(\omega)| = D(A, B). \end{aligned}$$

By Proposition 5.23 axioms IF-Div.3 and IF-Div.4 are equivalent. Let us see an example where axiom IF-Div.4 is violated. Consider $\Omega = \{\omega\}$ and the IF-sets:

$$A = \{(\omega, 0.25, 0.75)\} \text{ and } B = \{(\omega, 0, 0.5)\}.$$

It holds that $D_C(A, B) = 0$. If we consider $C = \{(\omega, 0.2, 0.6)\}$, it holds that:

$$A \cup C = \{\omega, 0.25, 0.6\} \text{ and } B \cup C = \{\omega, 0.2, 0.5\},$$

whence

$$D_C(A \cup C, B \cup C) = 0.025 > 0 = D_C(A, B).$$

Thus, D_C is neither an IF-divergence nor a dissimilitude.

In [89], Hong provided an example that showed that this IF-dissimilarity is a counterintuitive measure of comparison of fuzzy sets. The main reason is that:

$$\mu_A(\omega) - \nu_A(\omega) = \mu_B(\omega) - \nu_B(\omega) \quad \forall \omega \in \Omega \Rightarrow D_C(A, B) = 0.$$

In fact, if we consider the IF-sets A and B defined by:

$$A = \{(\omega, 0, 0) \mid \omega \in \Omega\} \text{ and } B = \{(\omega, 0.5, 0.5) \mid \omega \in \Omega\};$$

we obtain $D_C(A, B) = 0$. However, these IF-sets do not seem to be very similar.

Dengfenf & Chuntian dissimilarity Dengfenf and Chuntian ([111]) proposed the following IF-dissimilarity:

$$D_{DC}(A, B) = \frac{1}{\sqrt[p]{n}} \left(\sum_{\omega \in \Omega} \left| \frac{1}{2} (\mu_A(\omega) - \mu_B(\omega) - \nu_A(\omega) + \nu_B(\omega)) \right|^p \right),$$

for some $p \geq 1$. Again, it obviously holds that $D(A^c, B^c) = D(A, B)$, that is, D_{DC} satisfies IF-Div.5, and therefore, by Proposition 5.23, axioms IF-Div.3 and IF-Div.4 are equivalent. Furthermore, when $p = 1$, D_{DC} becomes Chen dissimilarity multiplied by a constant. Thus, in order to obtain a counterexample, it suffices to consider the same than in the previous paragraph.

Hung & Yang dissimilarities Previously we have seen some examples of IF-dissimilarities proposed by Hung and Yang that are also IF-divergences. Here we give some examples of IF-dissimilarities proposed by them which are not IF-divergences. They are given by:

$$D_{\omega 1}(A, B) = 1 - \frac{1}{n} \sum_{\omega \in \Omega} \frac{\min(\mu_A(\omega), \mu_B(\omega)) + \min(\nu_A(\omega), \nu_B(\omega))}{\max(\mu_A(\omega), \mu_B(\omega)) + \max(\nu_A(\omega), \nu_B(\omega))}.$$

$$D_{pk1}(A, B) = 1 - \frac{\sum_{\omega \in \Omega} \min(\mu_A(\omega), \mu_B(\omega)) + \min(\nu_A(\omega), \nu_B(\omega))}{\sum_{\omega \in \Omega} \max(\mu_A(\omega), \mu_B(\omega)) + \max(\nu_A(\omega), \nu_B(\omega))}.$$

$$D_{pk3}(A, B) = \frac{\sum_{\omega \in \Omega} |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|}{\sum_{\omega \in \Omega} |\mu_A(\omega) + \mu_B(\omega)| + |\nu_A(\omega) + \nu_B(\omega)|}.$$

These dissimilarities satisfy axiom IF-Div.5, and therefore, using Proposition 5.23, both axioms IF-Div.3 and IF-Div.4 become equivalent. However, none of them satisfies these axioms. Let us give a counterexample for D_{ω_1} : consider an universe $\Omega = \{\omega\}$ and the IF-sets:

$$A = \{(\omega, 0.75, 0.19)\} \text{ and } B = \{(\omega, 0.48, 0.23)\}.$$

For these IF-sets, $D_{\omega_1}(A, B) = 0.32$. If we now consider the IF-set $C = \{(\omega, 0.25, 0.06)\}$, then $A \cup C$ and $B \cup C$ are given by:

$$A \cup C = \{(\omega, 0.75, 0.06)\} \text{ and } B \cup C = \{(\omega, 0.48, 0.06)\}.$$

Hence,

$$D_{\omega_1}(A \cup C, B \cup C) \geq 0.333 > 0.32 = D(A, B).$$

The same example shows that D_{pk1} does not satisfy IF-Div.4, since for $n = 1$ D_{pk1} and D_{ω_1} are the same function.

Let us prove now that D_{ω_3} does not satisfy IF-Div.4 neither. For this, take $\Omega = \{\omega\}$ define the following IF-sets:

$$A = \{(\omega, 0.24, 0.28)\}, B = \{(\omega, 0.66, 0.29)\}, C = \{(\omega, 0.02, 0.15)\}.$$

Then, it holds that:

$$D_{pk3}(A, B) = 0.29 < 0.35 = D_{pk3}(A \cup C, B \cup C).$$

Thus, none of these IF-dissimilarity measures are IF-divergences or IF-dissimilitudes.

In Table 5.1 we have summarized the results we have presented in this section. There, we can see which axioms satisfy every one of the examples of IF-dissimilarities we have studied. We can remark that all these examples satisfy the property IF-Div.5, and then IF-Div.3 and IF-Div.4 are equivalent. Recall that all the measures we have studied satisfy property IF-Div.5, and then IF-divergences and IF-dissimilitudes become equivalent.

5.1.4 Local IF-divergences

In this section we are going to study a special type of IF-divergences called the local IF-divergences. They are an important family of IF-divergences because of the interesting properties they satisfy.

Let us consider a universe $\Omega = \{\omega_1, \dots, \omega_n\}$ and an IF-divergence D_{IFS} defined on $IFSS(\Omega) \times IFSS(\Omega)$. From IF-Div.4, we know that $D(A \cup C, B \cup C) \leq D(A, B)$ for every $C \in IFSS(\Omega)$. In particular, given $C = \{\omega_i\}$, we can express it equivalently by

$$C = \{(\omega_i, 1, 0), (\omega_j, 0, 1) \mid j \neq i\}.$$

Name	Notation	IF-Diss.1&2	IF-Div.3&4	IF-Div.5	IF-diss	IF-div
Hamming	l_{IFS}	OK	OK	OK	Yes	Yes
Normalized Hamming	l_{nIFS}	OK	OK	OK	Yes	Yes
Hausdorff	d_{H}	OK	OK	OK	Yes	Yes
Normalized Hausdorff	d_{nH}	OK	OK	OK	Yes	Yes
Normalized Eucliden	q_{IFS}	OK	FAIL	OK	Yes	No
Hong and Kim (I)	D_{C}	OK	OK	OK	Yes	Yes
Hong and Kim (II)	D_{L}	OK	OK	OK	Yes	Yes
Li et al.	D_{O}	OK	OK	OK	Yes	Yes
Mitchell	D_{HB}	OK	OK	OK	Yes	Yes
Liang and Shi (I)	D_{e}^p	OK	OK	OK	Yes	Yes
Liang and Shi (II)	D_{h}^p	OK	OK	OK	Yes	Yes
Liang and Shi (III)	D_{s}^p	OK	FAIL	OK	Yes	No
Chen	D_{C}	OK	FAIL	OK	Yes	No
Dengfeng and Chuntian	D_{DC}	OK	FAIL	OK	Yes	No
Hung and Yang (I)	D_{HY}^1	OK	OK	OK	Yes	Yes
Hung and Yang (II)	D_{HY}^2	OK	OK	OK	Yes	Yes
Hung and Yang (III)	D_{HY}^3	OK	OK	OK	Yes	Yes
Hung and Yang (IV)	$D_{\omega 1}$	OK	FAIL	OK	Yes	No
Hung and Yang (V)	$D_{\text{pk}1}$	OK	FAIL	OK	Yes	No
Hung and Yang (VI)	$D_{\text{pk}2}$	OK	OK	OK	Yes	Yes
Hung and Yang (VIII)	$D_{\text{pk}3}$	OK	FAIL	OK	Yes	No

Table 5.1: Behaviour of well-known dissimilarities and IF-divergences.

Then, the IF-sets $A \cup \{\omega_i\}$ and $B \cup \{\omega_i\}$ are given by:

$$\begin{aligned} A \cup \{\omega_i\} &= \{(\omega_i, 1, 0), (\omega_j, \mu_A(\omega_j), \nu_A(\omega_j)) \mid j \neq i\}. \\ B \cup \{\omega_i\} &= \{(\omega_i, 1, 0), (\omega_j, \mu_B(\omega_j), \nu_B(\omega_j)) \mid j \neq i\}. \end{aligned}$$

Applying axiom IF-Div.4 to these IF-sets, we obtain the following inequality:

$$D_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = D_{\text{IFS}}(A, B).$$

Hence, the only difference between $D_{\text{IFS}}(A \cup C, B \cup C)$ and $D_{\text{IFS}}(A, B)$ is on the i -th element. However, such a function may not exist. When it does, the IF-divergence will be called local.

Definition 5.28 Let D_{IFS} be an IF-divergence. It is called local (or it is said to satisfy the local property) when for every $A, B \in \text{IFSS}(\Omega)$ and every $\omega \in \Omega$ it holds that:

$$D_{\text{IFS}}(A, B) - D_{\text{IFS}}(A \cup \{\omega\}, B \cup \{\omega\}) = h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)). \quad (5.3)$$

In order to characterize local IF-divergences we are going to see the next Theorem.

Theorem 5.29 A map $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ on a finite universe $\Omega = \{\omega_1, \dots, \omega_n\}$ is a local IF-divergence if and only if there is a function $h_{\text{IFS}} : \mathcal{T}^2 \rightarrow \mathbb{R}$ such that for every $A, B \in \text{IFSS}(\Omega)$:

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)), \quad (5.4)$$

where \mathcal{T} denotes the set $\mathcal{T} = \{(t, z) \in [0, 1]^2 \mid t + z \leq 1\}$ and h_{IFS} fulfills the following properties:

- IF-loc.1 $h_{\text{IFS}}(x, y, x, y) = 0$ for every $(x, y) \in \mathcal{T}$.
- IF-loc.2 $h_{\text{IFS}}(x_1, x_2, y_1, y_2) = h_{\text{IFS}}(y_1, y_2, x_1, x_2)$ for every $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$.
- IF-loc.3 If $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$, $z \in [0, 1]$ and $x_1 \leq z \leq y_1$, it holds that:
 $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, x_2, z, y_2)$.
 Moreover, if $(x_2, z), (y_2, z) \in \mathcal{T}$ it holds that
 $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(z, x_2, y_1, y_2)$.
- IF-loc.4 If $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$, $z \in [0, 1]$ and $x_2 \leq z \leq y_2$, it holds that:
 $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, x_2, y_1, z)$.
 Moreover, if $(x_1, z), (y_1, z) \in \mathcal{T}$ it holds that:
 $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, z, y_1, y_2)$.
- IF-loc.5 If $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$, then:
 $h_{\text{IFS}}(z, x_2, z, y_2) \leq h_{\text{IFS}}(x_1, x_2, y_1, y_2)$ if $(x_2, z), (y_2, z) \in \mathcal{T}$ and
 $h_{\text{IFS}}(x_1, z, y_1, z) \leq h_{\text{IFS}}(x_1, x_2, y_1, y_2)$ if $(x_1, z), (y_1, z) \in \mathcal{T}$.

Proof: Assume first of all that D_{IFS} is a local IF-divergence and let us prove that $D_{\text{IFS}}(A, B)$ can be expressed as in Equation (5.4) for every $A, B \in \text{IFSS}(\Omega)$, where h_{IFS} satisfies the properties IF-loc.1 to IF-loc.6. In order to prove that, we will apply recursively Equation (5.3):

$$\begin{aligned} D_{\text{IFS}}(A, B) &= D_{\text{IFS}}(A \cup \{\omega_1\}, B \cup \{\omega_1\}) \\ &\quad + h_{\text{IFS}}(\mu_A(\omega_1), \nu_A(\omega_1), \mu_B(\omega_1), \nu_B(\omega_1)) \\ &= D_{\text{IFS}}(A \cup \{\omega_1\} \cup \{\omega_2\}, B \cup \{\omega_1\} \cup \{\omega_2\}) \\ &\quad + \sum_{i=1}^2 h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)) \\ &= \dots \\ &= D_{\text{IFS}}(\Omega, \Omega) + \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)). \end{aligned}$$

Moreover, from axiom IF-Diss.1 we know that $D_{\text{IFS}}(\Omega, \Omega) = 0$, and therefore D_{IFS} can be expressed by:

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)).$$

This shows that D_{IFS} can be expressed as in Equation (5.4).

Let us prove next that h_{IFS} fulfills properties IF-loc.1 to IF-loc.5:

IF-loc.1: Take $x, y \in \mathcal{T}$, and let us prove that $h_{\text{IFS}}(x, y, x, y) = 0$. Define the IF-set A by $\mu_A(\omega_i) = x$ and $\nu_A(\omega_i) = y$, for every $i = 1, \dots, n$. Note that A is in fact an IF-set since $\mu_A(\omega_i) + \nu_A(\omega_i) = x + y \leq 1$ for every $i = 1, \dots, n$. Applying IF-diss.1, $D_{\text{IFS}}(A, A) = 0$, and therefore, since $D_{\text{IFS}}(A, A)$ can be expressed as in Equation (5.4), it holds that:

$$\begin{aligned} 0 = D_{\text{IFS}}(A, A) &= \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_A(\omega_i), \nu_A(\omega_i)) \\ &= \sum_{i=1}^n h_{\text{IFS}}(x, y, x, y) = n \cdot h_{\text{IFS}}(x, y, x, y). \end{aligned}$$

Then, it must hold that $h_{\text{IFS}}(x, y, x, y) = 0$.

IF-loc.2: Let $(x_1, x_2), (y_1, y_2)$ be two elements in \mathcal{T} . Consider the IF-sets A and B defined by: $\mu_A(\omega_i) = x_1$, $\nu_A(\omega_i) = x_2$, $\mu_B(\omega_i) = y_1$ and $\nu_B(\omega_i) = y_2$. Using axiom

IF-diss.2 and Equation (5.4) we obtain the following:

$$\begin{aligned}
n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) &= \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)) \\
&= D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A) \\
&= \sum_{i=1}^n h_{\text{IFS}}(\mu_B(\omega_i), \nu_B(\omega_i), \mu_A(\omega_i), \nu_A(\omega_i)) \\
&= n h_{\text{IFS}}(y_1, y_2, x_1, x_2).
\end{aligned}$$

Thus, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) = h_{\text{IFS}}(y_1, y_2, x_1, x_2)$.

IF-loc.3: Consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$ such that $x_1 \leq z \leq y_1$, and let us define the IF-sets A and B by: $\mu_A(\omega_i) = x_1, \nu_A(\omega_i) = x_2, \mu_B(\omega_i) = y_1$ and $\nu_B(\omega_i) = y_2$, for every $i = 1, \dots, n$. We have to consider two cases:

- On one hand we are going to prove that

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, x_2, z, y_2).$$

To see this, consider the IF-set C defined by $\mu_C(\omega_i) = z$ and $\nu_C(\omega_i) = 0$ for $i = 1, \dots, n$. Then the IF-sets $A \cap C$ and $B \cap C$ are given by:

$$\begin{aligned}
A \cap C &= A. \\
B \cap C &= \{(\omega_i, \mu_C(\omega_i), \nu_B(\omega_i)) \mid i = 1, \dots, n\}.
\end{aligned}$$

By axiom IF-Div.3, we see that $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cap C, B \cap C) = D_{\text{IFS}}(A, B \cap C)$, and then Equation (5.4) implies that:

$$\begin{aligned}
n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) &= D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cap C, B \cap C) \\
&= n \cdot h_{\text{IFS}}(x_1, x_2, z, y_2).
\end{aligned}$$

Hence, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, x_2, z, y_2)$.

- Let us prove now that, when $(x_2, z), (y_2, z) \in \mathcal{T}$, it holds that $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(z, x_2, y_1, y_2)$. Consider the IF-set C defined by $\mu_C(\omega_i) = z$ and $\nu_C(\omega_i) = \max(x_2, y_2)$, for $i = 1, \dots, n$. Note that C is an IF-set because $\mu_C(\omega_i) + \nu_C(\omega_i) = \max(x_2 + z, y_2 + z) \leq 1$, for $i = 1, \dots, n$. Using axiom IF-Div.4, we deduce that $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cup C, B \cup C)$. Moreover, the IF-sets $A \cup C$ and $B \cup C$ are given by:

$$\begin{aligned}
A \cup C &= \{(\omega_i, \mu_C(\omega_i), \nu_A(\omega_i)) \mid i = 1, \dots, n\}. \\
B \cup C &= B.
\end{aligned}$$

Then, $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cup C, B)$. This, together with Equation (5.4), implies that:

$$n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) = D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cup C, B \cup C) = n \cdot h_{\text{IFS}}(z, x_2, y_1, y_2).$$

Hence, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(z, x_2, y_1, y_2)$.

IF-loc.4: The proof is similar to that of IF-loc.3. Consider (x_1, x_2) and (y_1, y_2) in \mathcal{T} , and let z be a point in $[0, 1]$ such that $x_2 \leq z \leq y_2$. Define the IF-sets A and B by:

$$A = \{(\omega, x_1, x_2) \mid \omega \in \Omega\} \text{ and } B = \{(\omega, y_1, y_2) \mid \omega \in \Omega\}.$$

If we consider the IF-set C given by:

$$C = \{(\omega, 0, z) \mid \omega \in \Omega\},$$

then, the IF-sets $A \cup C$ and $B \cup C$ are given by:

$$A \cup C = A \text{ and } B \cup C = \{(\omega, y_1, z)\}.$$

Applying axiom IF-Div.4 we deduce that

$$D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cup C, B \cup C) = D_{\text{IFS}}(A, B \cup C),$$

and using now Equation (5.4), we obtain:

$$n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) = D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cup C, B \cup C) = n \cdot h_{\text{IFS}}(x_1, x_2, y_1, z).$$

Moreover, if $(x_1, z), (y_1, z) \in \mathcal{T}$, we consider the set

$$C' = \{(\omega, \max(x_1, y_1), z) \mid \omega \in \Omega\}.$$

Since $(x_1, z), (y_1, z) \in \mathcal{T}$, C' is an IF-set. Moreover, $A \cap C'$ and $B \cap C'$ are given by:

$$A \cap C' = \{(\omega, x_1, z) \mid \omega \in \Omega\} \text{ and } B \cap C' = B.$$

Using axiom IF-Div.3, we deduce that

$$D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cap C', B \cap C') = D_{\text{IFS}}(A \cap C', B),$$

and applying Equation (5.4),

$$n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) = D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(A \cap C', B \cap C') = n \cdot h_{\text{IFS}}(x_1, z, y_1, y_2).$$

Hence, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, z, y_1, y_2)$.

IF-loc.5: Let us consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$.

If we assume that $(x_2, z), (y_2, z) \in \mathcal{T}$, then $\max(x_2, y_2) + z \leq 1$; we consider the IF-sets A, B, C and D given by:

$$\begin{aligned} A &= \{(\omega, x_1, x_2) \mid \omega \in \Omega\}, & B &= \{(\omega, y_1, y_2) \mid \omega \in \Omega\}. \\ C &= \{(\omega, z, x_2) \mid \omega \in \Omega\}, & D &= \{(\omega, z, y_2) \mid \omega \in \Omega\}. \end{aligned}$$

From Proposition 5.25, we know that $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(C, D)$, and applying Equation (5.4) we deduce that

$$n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) = D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(C, D) = n \cdot h_{\text{IFS}}(z, x_2, z, y_2).$$

Thus, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(z, x_2, z, y_2)$.

If we assume now that $(x_1, z), (y_1, z) \in \mathcal{T}$, it holds that $\max(x_1, y_1) + z \leq 1$; we consider the IF-sets:

$$\begin{aligned} A &= \{(\omega, x_1, x_2) \mid \omega \in \Omega\}, & B &= \{(\omega, y_1, y_2) \mid \omega \in \Omega\}. \\ C &= \{(\omega, x_1, z) \mid \omega \in \Omega\}, & D &= \{(\omega, y_1, z) \mid \omega \in \Omega\}. \end{aligned}$$

Applying Corollary 5.26, $D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(C, D)$. Using Equation (5.4), we obtain:

$$n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) = D_{\text{IFS}}(A, B) \geq D_{\text{IFS}}(C, D) = n \cdot h_{\text{IFS}}(x_1, z, y_1, z).$$

Thus, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) \geq h_{\text{IFS}}(x_1, z, y_1, z)$.

Summarizing, if D_{IFS} is a local IF-divergence, then $D_{\text{IFS}}(A, B)$ can be expressed as in Equation (5.4) where the function h_{IFS} satisfies IF-loc.1 to IF-loc.5.

Let us prove the converse: that if a function D_{IFS} is defined by Equation (5.4), where h_{IFS} fulfills properties IF-loc.1 to IF-loc.5, then D_{IFS} is a local IF-divergence.

First of all, let us prove that D_{IFS} is an IF-divergence, i.e., that it satisfies axioms IF-Diss.1, IF-Diss.2, IF-Div.3 and IF-Div.4.

IF-Diss.1: Let A be an IF-set. Then, $D_{\text{IFS}}(A, A) = 0$ because

$$D_{\text{IFS}}(A, A) = \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_A(\omega_i), \nu_A(\omega_i)) = 0,$$

since IF-loc.1 implies that $h_{\text{IFS}}(x, y, x, y) = 0$ for every $(x, y) \in \mathcal{T}$, and in particular $(\mu_A(\omega_i), \nu_A(\omega_i)) \in \mathcal{T}$.

IF-Diss.2: Let A, B be IF-sets, and let us prove that $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A)$. By IF-loc.2, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) = h_{\text{IFS}}(y_1, y_2, x_1, x_2)$ for every $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$, as $(\mu_A(\omega_i), \nu_A(\omega_i)), (\mu_B(\omega_i), \nu_B(\omega_i)) \in \mathcal{T}$, whence

$$D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A).$$

IF-Div.3 & IF-Div.4: Consider three IF-sets A, B and C , and let us show that $D_{\text{IFS}}(A, B) \geq \max(D_{\text{IFS}}(A \cup C, B \cup C), D_{\text{IFS}}(A \cap C, B \cap C))$. Consider the following partition of Ω :

$$\begin{aligned} P_1 &= \{\omega \in \Omega \mid \max(\mu_A(\omega), \mu_B(\omega)) \leq \mu_C(\omega)\}. \\ P_2 &= \{\omega \in \Omega \mid \mu_A(\omega) \leq \mu_C(\omega) < \mu_B(\omega)\}. \\ P_3 &= \{\omega \in \Omega \mid \mu_B(\omega) \leq \mu_C(\omega) < \mu_A(\omega)\}. \\ P_4 &= \{\omega \in \Omega \mid \mu_C(\omega) < \min(\mu_A(\omega), \mu_B(\omega))\}. \\ Q_1 &= \{\omega \in \Omega \mid \max(\nu_A(\omega), \nu_B(\omega)) \leq \nu_C(\omega)\}. \\ Q_2 &= \{\omega \in \Omega \mid \nu_A(\omega) \leq \nu_C(\omega) < \nu_B(\omega)\}. \\ Q_3 &= \{\omega \in \Omega \mid \nu_B(\omega) \leq \nu_C(\omega) < \nu_A(\omega)\}. \\ Q_4 &= \{\omega \in \Omega \mid \nu_C(\omega) < \min(\nu_A(\omega), \nu_B(\omega))\}. \end{aligned}$$

Thus, $\Omega = \bigcup_{i=1}^4 \bigcup_{j=1}^4 (P_i \cap Q_j)$. We are going to prove that, for every $i, j \in \{1, \dots, 4\}$, if $\omega \in P_i \cap Q_j$ then both:

$$\begin{aligned} &h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)) \text{ and} \\ &h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)) \end{aligned}$$

are smaller than

$$h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)).$$

1. $\omega \in P_1 \cap Q_1$; by hypothesis, we have that:

$$\max(\mu_A(\omega), \mu_B(\omega)) \leq \mu_C(\omega) \text{ and } \max(\nu_A(\omega), \nu_B(\omega)) \leq \nu_C(\omega),$$

whence

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_C(\omega), & \nu_{A \cup C}(\omega) &= \nu_A(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_A(\omega), & \nu_{A \cap C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_C(\omega), & \nu_{B \cup C}(\omega) &= \nu_B(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_B(\omega), & \nu_{B \cap C}(\omega) &= \nu_C(\omega). \end{aligned}$$

Moreover, property IF-loc.5 can be applied since

$$\max(\nu_A(\omega), \nu_B(\omega)) + \mu_C(\omega) \leq \nu_C(\omega) + \mu_C(\omega) \leq 1,$$

whence $(\nu_A(\omega), \mu_C(\omega)), (\nu_B(\omega), \mu_C(\omega)) \in \mathcal{T}$ and therefore

$$\begin{aligned} &h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_C(\omega), \nu_A(\omega), \mu_C(\omega), \nu_B(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

Similarly,

$$\begin{aligned} &h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_B(\omega), \nu_C(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

Let us remark that, in the rest of the proof, axioms IF-loc.3, IF-loc.4 and IF-loc.5 are applicable because the previous hypotheses are satisfied.

2. $\omega \in P_1 \cap Q_2$; by hypothesis it holds that:

$$\mu_A(\omega), \mu_B(\omega) \leq \mu_C(\omega) \text{ and } \nu_A(\omega) \leq \nu_C(\omega) < \nu_B(\omega),$$

whence

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_C(\omega), & \nu_{A \cup C}(\omega) &= \nu_A(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_A(\omega), & \nu_{A \cap C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_C(\omega), & \nu_{B \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_B(\omega), & \nu_{B \cap C}(\omega) &= \nu_B(\omega), \end{aligned}$$

As a consequence, by IF-loc.4 and IF-loc.5:

$$\begin{aligned} & h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ & \geq h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_C(\omega)) \\ & \geq h_{\text{IFS}}(\mu_C(\omega), \nu_A(\omega), \mu_C(\omega), \nu_C(\omega)) \\ & = h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

Similarly, by IF-loc.4:

$$\begin{aligned} & h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ & \geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_B(\omega), \nu_B(\omega)) \\ & = h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

3. $\omega \in P_1 \cap Q_3$; this case is immediate from case 2, if we exchange the roles of A and B .

4. $\omega \in P_1 \cap Q_4$; then we know that:

$$\mu_A(\omega), \mu_B(\omega) \leq \mu_C(\omega), \text{ and } \nu_C(\omega) < \nu_A(\omega), \nu_B(\omega).$$

Then, it holds that $A \cup C = B \cup C = C$, $A \cap C = A$ and $B \cap C = B$, whence

$$h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) = h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)).$$

Moreover,

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) & \geq 0 = h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_C(\omega), \nu_C(\omega)) \\ & = h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

5. $\omega \in P_2 \cap Q_1$; in that case we know that:

$$\mu_A(\omega) \leq \mu_C(\omega) < \mu_B(\omega) \text{ and } \nu_A(\omega), \nu_B(\omega) \leq \nu_C(\omega),$$

whence

$$\begin{aligned} \mu_{A \cup C}(\omega) & = \mu_C(\omega), & \nu_{A \cup C}(\omega) & = \nu_A(\omega), \\ \mu_{A \cap C}(\omega) & = \mu_A(\omega), & \nu_{A \cap C}(\omega) & = \nu_C(\omega), \\ \mu_{B \cup C}(\omega) & = \mu_B(\omega), & \nu_{B \cup C}(\omega) & = \nu_B(\omega), \\ \mu_{B \cap C}(\omega) & = \mu_C(\omega), & \nu_{B \cap C}(\omega) & = \nu_C(\omega). \end{aligned}$$

Thus, by IF-loc.3,

$$\begin{aligned} & h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ & \geq h_{\text{IFS}}(\mu_C(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ & = h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

Similarly, by IF-loc.1 and IF-loc.3,

$$\begin{aligned} & h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ & \geq 0 = h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_C(\omega), \nu_C(\omega)) \\ & \geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_C(\omega), \nu_C(\omega)) \\ & = h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

6. $\omega \in P_2 \cap Q_2$; we know that:

$$\mu_A(\omega) \leq \mu_C(\omega) < \mu_B(\omega) \text{ and } \nu_A(\omega) \leq \nu_C(\omega) < \nu_B(\omega).$$

Then

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_C(\omega), & \nu_{A \cup C}(\omega) &= \nu_A(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_A(\omega), & \nu_{A \cap C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_B(\omega), & \nu_{B \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_C(\omega), & \nu_{B \cap C}(\omega) &= \nu_B(\omega), \end{aligned}$$

and therefore, by IF-loc.3 and IF-loc.4,

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_C(\omega)) \\ &\geq h_{\text{IFS}}(\mu_C(\omega), \nu_A(\omega), \mu_B(\omega), \nu_C(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

As a consequence,

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_C(\omega), \nu_B(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

7. $\omega \in P_2 \cap Q_3$; we know that:

$$\mu_A(\omega) \leq \mu_C(\omega) < \mu_B(\omega) \text{ and } \nu_B(\omega) \leq \nu_C(\omega) < \nu_A(\omega).$$

Thus,

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_C(\omega), & \nu_{A \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_A(\omega), & \nu_{A \cap C}(\omega) &= \nu_A(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_B(\omega), & \nu_{B \cup C}(\omega) &= \nu_B(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_C(\omega), & \nu_{B \cap C}(\omega) &= \nu_C(\omega) \end{aligned}$$

whence, applying IF-loc.3 and IF-loc.4,

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)), \end{aligned}$$

and as a consequence

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_C(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_C(\omega), \nu_C(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

8. $\omega \in P_2 \cap Q_4$; it holds that:

$$\mu_A(\omega) \leq \mu_C(\omega) < \mu_B(\omega) \text{ and } \nu_C(\omega) \leq \nu_A(\omega), \nu_B(\omega),$$

whence

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_C(\omega), & \nu_{A \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_A(\omega), & \nu_{A \cap C}(\omega) &= \nu_A(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_B(\omega), & \nu_{B \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_C(\omega), & \nu_{B \cap C}(\omega) &= \nu_B(\omega). \end{aligned}$$

and thus, by IF-loc.3,

$$\begin{aligned} &h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &= h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_C(\omega), \nu_C(\omega)) \\ &\geq h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_B(\omega), \nu_C(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

In addition,

$$\begin{aligned} &h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &\geq h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_C(\omega), \nu_B(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

9. $\omega \in P_3 \cap Q_i$; this case is immediate if we exchange the roles of A and B and apply the case when $\omega \in P_2 \cap Q_i$.

10. $\omega \in P_4 \cap Q_1$; in such case

$$\mu_C(\omega) < \mu_A(\omega), \mu_B(\omega) \text{ and } \nu_A(\omega), \nu_B(\omega) \leq \nu_C(\omega).$$

We have that:

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_A(\omega), & \nu_{A \cup C}(\omega) &= \nu_A(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_C(\omega), & \nu_{A \cap C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_B(\omega), & \nu_{B \cup C}(\omega) &= \nu_B(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_C(\omega), & \nu_{B \cap C}(\omega) &= \nu_C(\omega), \end{aligned}$$

whence

$$h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) = h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)),$$

and moreover, by IF-loc.1,

$$\begin{aligned} &h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ &= 0 \geq h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_C(\omega), \nu_C(\omega)) \\ &= h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

11. $\omega \in P_4 \cap Q_2$; in such case we know that

$$\mu_C(\omega) < \mu_A(\omega), \mu_B(\omega) \text{ and } \nu_A(\omega) \leq \nu_C(\omega) < \nu_B(\omega).$$

It holds that

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_A(\omega), & \nu_{A \cup C}(\omega) &= \nu_A(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_C(\omega), & \nu_{A \cap C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_B(\omega), & \nu_{B \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_C(\omega), & \nu_{B \cap C}(\omega) &= \nu_B(\omega), \end{aligned}$$

whence, applying IF-loc.4,

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ \geq h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_C(\omega)) \\ = h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)). \end{aligned}$$

Moreover, applying IF-loc.1 and IF-loc.4,

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ = 0 \geq h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_C(\omega), \nu_C(\omega)) \\ \geq h_{\text{IFS}}(\mu_C(\omega), \nu_C(\omega), \mu_C(\omega), \nu_B(\omega)) \\ = h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

12. $\omega \in P_4 \cap Q_3$; this follows from the previous case by exchanging the roles of A and B .

13. $\omega \in P_4 \cap Q_4$; we know that

$$\mu_C(\omega) < \mu_A(\omega), \mu_B(\omega) \text{ and } \nu_C(\omega) < \nu_A(\omega), \nu_B(\omega)$$

whence

$$\begin{aligned} \mu_{A \cup C}(\omega) &= \mu_A(\omega), & \nu_{A \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{A \cap C}(\omega) &= \mu_C(\omega), & \nu_{A \cap C}(\omega) &= \nu_A(\omega), \\ \mu_{B \cup C}(\omega) &= \mu_B(\omega), & \nu_{B \cup C}(\omega) &= \nu_C(\omega), \\ \mu_{B \cap C}(\omega) &= \mu_C(\omega), & \nu_{B \cap C}(\omega) &= \nu_B(\omega), \end{aligned}$$

and thus by IF-loc.5

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ \geq h_{\text{IFS}}(\mu_A(\omega), \nu_C(\omega), \mu_B(\omega), \nu_C(\omega)) \\ = h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

Moreover:

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) \\ \geq h_{\text{IFS}}(\mu_C(\omega), \nu_A(\omega), \mu_C(\omega), \nu_B(\omega)) \\ = h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega)). \end{aligned}$$

Hence, since $\Omega = \bigcup_{i=1}^4 \bigcup_{j=1}^4 P_i \cap Q_j$, we conclude that for all $\omega \in \Omega$ it holds that:

$$\begin{aligned} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) &\geq \\ &\max(h_{\text{IFS}}(\mu_{A \cup C}(\omega), \nu_{A \cup C}(\omega), \mu_{B \cup C}(\omega), \nu_{B \cup C}(\omega)), \\ &h_{\text{IFS}}(\mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega), \mu_{B \cap C}(\omega), \nu_{B \cap C}(\omega))). \end{aligned}$$

Thus, D_{IFS} satisfies both IF-Div.3 and IF-Div.4, and therefore it is an IF-divergence. It only remains to show that D_{IFS} is local. But this holds trivially, taking into account that

$$\begin{aligned} D_{\text{IFS}}(A, B) - D_{\text{IFS}}(A \cup \{\omega_j\}, B \cup \{\omega_j\}) &= \\ &= \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)) \\ &- \sum_{i \neq j} h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)) - h_{\text{IFS}}(1, 1, 0, 0) \\ &= h_{\text{IFS}}(\mu_A(\omega_j), \nu_A(\omega_j), \mu_B(\omega_j), \nu_B(\omega_j)). \end{aligned}$$

We conclude that D_{IFS} is a local IF-divergence. ■

Properties of local IF-divergences

In this section we are going to study some properties of local IF-divergences. In some cases, the local property will allow us to obtain interesting and useful properties.

We begin by studying under which conditions a local divergence satisfies IF-Div.5.

Proposition 5.30 *Let D_{IFS} be a local IF-divergence which associated function h_{IFS} . It satisfies IF-Div.5 if and only if for every*

$$(x_1, x_2), (y_1, y_2) \in \mathcal{T} = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}$$

it holds that

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = h_{\text{IFS}}(x_2, x_1, y_2, y_1).$$

Proof: Assume that D_{IFS} satisfies axiom IF-Div.5, i.e., that for every $A, B \in \text{IFSs}(\Omega)$, $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c)$. Consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$, and define the IF-sets A and B by:

$$A = \{(\omega, x_1, x_2) \mid \omega \in \Omega\} \text{ and } B = \{(\omega, y_1, y_2) \mid \omega \in \Omega\}.$$

By IF-Div.5, it holds that $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c)$. Using Equation (5.4),

$$n \cdot h_{\text{IFS}}(x_1, x_2, y_1, y_2) = D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c) = n \cdot h_{\text{IFS}}(x_2, x_1, y_2, y_1).$$

Thus, $h_{\text{IFS}}(x_1, x_2, y_1, y_2) = h_{\text{IFS}}(x_2, x_1, y_2, y_1)$.

Conversely, assume that $h_{\text{IFS}}(x_1, x_2, y_1, y_2) = h_{\text{IFS}}(x_2, x_1, y_2, y_1)$ for every two elements $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$. Let A and B be two IF-sets. Then, for every $i = 1, \dots, n$ it holds that:

$$h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)) = h_{\text{IFS}}(\nu_A(\omega_i), \mu_A(\omega_i), \nu_B(\omega_i), \mu_B(\omega_i))$$

and therefore $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c)$. ■

Next we give a lemma that shall be useful later.

Lemma 5.31 *If D_{IFS} is a local IF-divergence, then for every $i = 1, \dots, n$ it holds that*

$$D_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = D_{\text{IFS}}(A \cap \{\omega_i\}^c, B \cap \{\omega_i\}^c).$$

Proof: Consider the IF-sets $A \cap \{\omega_i\}^c$ and $B \cap \{\omega_i\}^c$. Note that

$$\begin{aligned} (A \cap \{\omega_i\}^c) \cup \{\omega_i\} &= (A \cup \{\omega_i\}) \cap (\{\omega_i\}^c \cup \{\omega_i\}) = A \cup \{\omega_i\}. \\ (B \cap \{\omega_i\}^c) \cup \{\omega_i\} &= (B \cup \{\omega_i\}) \cap (\{\omega_i\}^c \cup \{\omega_i\}) = B \cup \{\omega_i\}. \end{aligned}$$

Since D_{IFS} is a local IF-divergence,

$$\begin{aligned} D_{\text{IFS}}(A \cap \{\omega_i\}^c, B \cap \{\omega_i\}^c) &- D_{\text{IFS}}((A \cap \{\omega_i\}^c) \cup \{\omega_i\}, (B \cap \{\omega_i\}^c) \cup \{\omega_i\}) \\ &= D_{\text{IFS}}(A \cap \{\omega_i\}^c, B \cap \{\omega_i\}^c) - D_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) \\ &= h_{\text{IFS}}(\mu_{A \cap \{\omega_i\}^c}(\omega_i), \nu_{A \cap \{\omega_i\}^c}(\omega_i), \mu_{B \cap \{\omega_i\}^c}(\omega_i), \nu_{B \cap \{\omega_i\}^c}(\omega_i)) \\ &= h_{\text{IFS}}(0, 1, 0, 1) = 0, \end{aligned}$$

using that

$$\begin{aligned} \mu_{A \cap \{\omega_i\}^c}(\omega_i) &= \min(\mu_A(\omega_i), 0) = 0, \\ \nu_{A \cap \{\omega_i\}^c}(\omega_i) &= \max(\nu_A(\omega_i), 1) = 1, \\ \mu_{B \cap \{\omega_i\}^c}(\omega_i) &= \min(\mu_B(\omega_i), 0) = 0, \\ \nu_{B \cap \{\omega_i\}^c}(\omega_i) &= \max(\nu_B(\omega_i), 1) = 1. \end{aligned} \quad \blacksquare$$

Using this lemma, we can establish the following proposition.

Proposition 5.32 *An IF-divergence D_{IFS} is local if and only if there is a function h' such that*

$$D_{\text{IFS}}(A, B) - D_{\text{IFS}}(A \cap \{\omega_i\}^c, B \cap \{\omega_i\}^c) = h'(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i))$$

for every $A, B \in \text{IFSs}(\Omega)$.

Proof: It is immediate from the previous lemma. ■

Let us give another characterization of local IF-divergences.

Proposition 5.33 *An IF-divergence D_{IFS} is local if and only if for every $X \in \mathcal{P}(\Omega)$ it holds that:*

$$D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A \cap X, B \cap X) + D_{\text{IFS}}(A \cap X^c, B \cap X^c),$$

for every $A, B \in \text{IFSS}(\Omega)$.

Proof: Assume that D_{IFS} is a local IF-divergence, and let us consider $A, B \in \text{IFSS}(\Omega)$ and $X \in \mathcal{P}(\Omega)$.

Since $A = (A \cap X) \cup (A \cap X^c)$ and $B = (B \cap X) \cup (B \cap X^c)$, it holds that

$$D_{\text{IFS}}(A, B) = D_{\text{IFS}}((A \cap X) \cup (A \cap X^c), (B \cap X) \cup (B \cap X^c)).$$

Taking into account that D_{IFS} is local, we deduce that:

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n h_{\text{IFS}}(\mu_{(A \cap X) \cup (A \cap X^c)}(\omega_i), \nu_{(A \cap X) \cup (A \cap X^c)}(\omega_i), \mu_{(B \cap X) \cup (B \cap X^c)}(\omega_i), \nu_{(B \cap X) \cup (B \cap X^c)}(\omega_i)).$$

Moreover, by splitting the sum between the elements on X and X^c ,

$$\begin{aligned} D_{\text{IFS}}(A, B) &= \sum_{\omega \in X} h_{\text{IFS}}(\mu_{(A \cap X) \cup (A \cap X^c)}(\omega), \nu_{(A \cap X) \cup (A \cap X^c)}(\omega), \mu_{(B \cap X) \cup (B \cap X^c)}(\omega), \nu_{(B \cap X) \cup (B \cap X^c)}(\omega)) \\ &\quad + \sum_{\omega \in X^c} h_{\text{IFS}}(\mu_{(A \cap X) \cup (A \cap X^c)}(\omega), \nu_{(A \cap X) \cup (A \cap X^c)}(\omega), \mu_{(B \cap X) \cup (B \cap X^c)}(\omega), \nu_{(B \cap X) \cup (B \cap X^c)}(\omega)). \end{aligned}$$

Furthermore:

$$\begin{aligned} \omega \in X &\Rightarrow \begin{cases} \mu_{(A \cap X) \cup (A \cap X^c)}(\omega) = \max(\mu_{A \cap X}(\omega), \mu_{A \cap X^c}(\omega)) \\ \quad = \max(\mu_{A \cap X}(\omega), 0) = \mu_{A \cap X}(\omega). \\ \nu_{(A \cap X) \cup (A \cap X^c)}(\omega) = \min(\nu_{A \cap X}(\omega), \nu_{A \cap X^c}(\omega)) \\ \quad = \min(\nu_{A \cap X}(\omega), 1) = \nu_{A \cap X}(\omega). \end{cases} \\ \omega \in X^c &\Rightarrow \begin{cases} \mu_{(A \cap X) \cup (A \cap X^c)}(\omega) = \max(\mu_{A \cap X}(\omega), \mu_{A \cap X^c}(\omega)) \\ \quad = \max(0, \mu_{A \cap X^c}(\omega)) = \mu_{A \cap X^c}(\omega). \\ \nu_{(A \cap X) \cup (A \cap X^c)}(\omega) = \min(\nu_{A \cap X}(\omega), \nu_{A \cap X^c}(\omega)) \\ \quad = \min(1, \nu_{A \cap X^c}(\omega)) = \nu_{A \cap X^c}(\omega). \end{cases} \end{aligned}$$

Similarly,

$$\begin{aligned} \omega \in X &\Rightarrow \begin{cases} \mu_{(B \cap X) \cup (B \cap X^c)}(\omega) = \mu_{B \cap X}(\omega). \\ \nu_{(B \cap X) \cup (B \cap X^c)}(\omega) = \nu_{B \cap X}(\omega). \end{cases} \\ \omega \in X^c &\Rightarrow \begin{cases} \mu_{(B \cap X) \cup (B \cap X^c)}(\omega) = \mu_{B \cap X^c}(\omega). \\ \nu_{(B \cap X) \cup (B \cap X^c)}(\omega) = \nu_{B \cap X^c}(\omega). \end{cases} \end{aligned}$$

Thus, the expression of $D_{\text{IFS}}(A, B)$ becomes

$$D_{\text{IFS}}(A, B) = \sum_{\omega \in X} h_{\text{IFS}}(\mu_{A \cap X}(\omega), \nu_{A \cap X}(\omega), \mu_{B \cap X}(\omega), \nu_{B \cap X}(\omega)) \\ + \sum_{\omega \in X^c} h_{\text{IFS}}(\mu_{A \cap X^c}(\omega), \nu_{A \cap X^c}(\omega), \mu_{B \cap X^c}(\omega), \nu_{B \cap X^c}(\omega)).$$

Taking into account that

$$D_{\text{IFS}}(A \cap X, B \cap X) = \sum_{\omega \in \Omega} h_{\text{IFS}}(\mu_{A \cap X}(\omega), \nu_{A \cap X}(\omega), \mu_{B \cap X}(\omega), \nu_{B \cap X}(\omega)) \\ = \sum_{\omega \in X} h_{\text{IFS}}(\mu_{A \cap X}(\omega), \nu_{A \cap X}(\omega), \mu_{B \cap X}(\omega), \nu_{B \cap X}(\omega)) \\ + \sum_{\omega \in X^c} h_{\text{IFS}}(\mu_{A \cap X}(\omega), \nu_{A \cap X}(\omega), \mu_{B \cap X}(\omega), \nu_{B \cap X}(\omega)) \\ = \sum_{\omega \in X} h_{\text{IFS}}(\mu_{A \cap X}(\omega), \nu_{A \cap X}(\omega), \mu_{B \cap X}(\omega), \nu_{B \cap X}(\omega)), \\ D_{\text{IFS}}(A \cap X^c, B \cap X^c) = \sum_{\omega \in \Omega} h_{\text{IFS}}(\mu_{A \cap X^c}(\omega), \nu_{A \cap X^c}(\omega), \mu_{B \cap X^c}(\omega), \nu_{B \cap X^c}(\omega)) \\ = \sum_{\omega \in X} h_{\text{IFS}}(\mu_{A \cap X^c}(\omega), \nu_{A \cap X^c}(\omega), \mu_{B \cap X^c}(\omega), \nu_{B \cap X^c}(\omega)) \\ + \sum_{\omega \in X^c} h_{\text{IFS}}(\mu_{A \cap X^c}(\omega), \nu_{A \cap X^c}(\omega), \mu_{B \cap X^c}(\omega), \nu_{B \cap X^c}(\omega)) \\ = \sum_{\omega \in X^c} h_{\text{IFS}}(\mu_{A \cap X^c}(\omega), \nu_{A \cap X^c}(\omega), \mu_{B \cap X^c}(\omega), \nu_{B \cap X^c}(\omega)),$$

we conclude that

$$D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A \cap X, B \cap X) + D_{\text{IFS}}(A \cap X^c, B \cap X^c).$$

Conversely, assume that $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A \cap X, B \cap X) + D_{\text{IFS}}(A \cap X^c, B \cap X^c)$ for every $A, B \in \text{IFSS}(\Omega)$ and $X \subseteq \Omega$. Applying this property to the crisp set $X = \{\omega_1\}$,

$$D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A \cap \{\omega_1\}, B \cap \{\omega_1\}) + D_{\text{IFS}}(A \cap \{\omega_2, \dots, \omega_n\}, B \cap \{\omega_2, \dots, \omega_n\}) \\ = D_{\text{IFS}}(A_1, B_1) + D_{\text{IFS}}(A \cap \{\omega_2, \dots, \omega_n\}, B \cap \{\omega_2, \dots, \omega_n\}),$$

where the IF-sets A_1 and B_1 are defined by

$$A_1 = \{(\omega_1, \mu_A(\omega_1), \nu_A(\omega_1)), (\omega_i, 0, 1) \mid i \neq 1\}, \\ B_1 = \{(\omega_1, \mu_B(\omega_1), \nu_B(\omega_1)), (\omega_i, 0, 1) \mid i \neq 1\}.$$

Now, apply the hypothesis to the crisp set $X = \{\omega_2\}$ and the IF-sets $A \cap \{\omega_2, \dots, \omega_n\}$ and $B \cap \{\omega_2, \dots, \omega_n\}$.

$$D_{\text{IFS}}(A \cap \{\omega_2, \dots, \omega_n\}, B \cap \{\omega_2, \dots, \omega_n\}) = D_{\text{IFS}}(A \cap \{\omega_2\}, B \cap \{\omega_2\}) \\ + D_{\text{IFS}}(A \cap \{\omega_3, \dots, \omega_n\}, B \cap \{\omega_3, \dots, \omega_n\}) \\ = D_{\text{IFS}}(A_2, B_2) \\ + D_{\text{IFS}}(A \cap \{\omega_3, \dots, \omega_n\}, B \cap \{\omega_3, \dots, \omega_n\}),$$

where

$$\begin{aligned} A_2 &= \{(\omega_2, \mu_A(\omega_2), \nu_A(\omega_2)), (\omega_i, 0, 1) \mid i \neq 2\}, \\ B_2 &= \{(\omega_2, \mu_B(\omega_2), \nu_B(\omega_2)), (\omega_i, 0, 1) \mid i \neq 2\}. \end{aligned}$$

If we repeat the process, for any $j \in \{1, \dots, n-1\}$, given $X = \{\omega_j\}$ and the IF-sets $A \cap \{\omega_j, \dots, \omega_n\}$ and $B \cap \{\omega_j, \dots, \omega_n\}$, it holds that:

$$\begin{aligned} D_{\text{IFS}}(A \cap \{\omega_j, \dots, \omega_n\}, B \cap \{\omega_j, \dots, \omega_n\}) \\ &= D_{\text{IFS}}(A \cap \{\omega_j\}, B \cap \{\omega_j\}) + D_{\text{IFS}}(A \cap \{\omega_{j+1}, \dots, \omega_n\}, B \cap \{\omega_{j+1}, \dots, \omega_n\}) \\ &= D_{\text{IFS}}(A_j, B_j) + D_{\text{IFS}}(A \cap \{\omega_{j+1}, \dots, \omega_n\}, B \cap \{\omega_{j+1}, \dots, \omega_n\}), \end{aligned}$$

where

$$\begin{aligned} A_j &= \{(\omega_j, \mu_A(\omega_j), \nu_A(\omega_j)), (\omega_i, 0, 1) \mid i \neq j\}, \\ B_j &= \{(\omega_j, \mu_B(\omega_j), \nu_B(\omega_j)), (\omega_i, 0, 1) \mid i \neq j\}. \end{aligned}$$

Then, $D_{\text{IFS}}(A, B)$ can be expressed by

$$\begin{aligned} D_{\text{IFS}}(A, B) &= D_{\text{IFS}}(A_1, B_1) + D_{\text{IFS}}(A \cap \{\omega_2, \dots, \omega_n\}, B \cap \{\omega_2, \dots, \omega_n\}) \\ &= D_{\text{IFS}}(A_1, B_1) + D_{\text{IFS}}(A_2, B_2) \\ &\quad + D_{\text{IFS}}(A \cap \{\omega_3, \dots, \omega_n\}, B \cap \{\omega_3, \dots, \omega_n\}) \\ &= \dots = \sum_{i=1}^n D_{\text{IFS}}(A_i, B_i). \end{aligned}$$

Now, consider the difference between $D_{\text{IFS}}(A, B)$ and $D_{\text{IFS}}(A \cap \{\omega_i\}, B \cap \{\omega_i\})$:

$$D_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) - D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A_i \cup \{\omega_i\}, B_i \cup \{\omega_i\}) - D_{\text{IFS}}(A_i, B_i).$$

This difference only depends on $\mu_A(\omega_i), \nu_A(\omega_i)$ and $\mu_B(\omega_i), \nu_B(\omega_i)$, so taking into account Definition 5.28 we conclude that D_{IFS} is a local IF-divergence. ■

A particular case of interest is the comparison of an IF-set and its complementary. In this sense, it seems useful to measure how imprecise an IF-set is. We consider the following partial order between IF-sets: given two IF-sets A and B , we say that A is sharper than B , and denote it $A \lll B$, when $|\mu_A(\omega) - 0.5| \geq |\mu_B(\omega) - 0.5|$ and $|\nu_A(\omega) - 0.5| \geq |\nu_B(\omega) - 0.5|$ for every $\omega \in \Omega$.

Using this partial order we can establish the following interesting property.

Proposition 5.34 *If D_{IFS} is a local IF-divergence and $A \lll B$, then it holds that $D_{\text{IFS}}(A, A^c) \geq D_{\text{IFS}}(B, B^c)$.*

Proof: Assume that $A \lll B$, and let us consider the crisp sets X and Y defined by

$$\begin{aligned} X &= \{\omega \in \Omega \mid \mu_A(\omega) \leq 0.5 \text{ and } \nu_A(\omega) \geq 0.5\}. \\ Y &= \{\omega \in \Omega \mid \mu_B(\omega) \leq 0.5 \text{ and } \nu_B(\omega) \geq 0.5\}. \end{aligned}$$

Applying Proposition 5.33,

$$D_{\text{IFS}}(A, A^c) = D_{\text{IFS}}(A \cap X, A^c \cap X) + D_{\text{IFS}}(A \cap X^c, A^c \cap X^c)$$

and if we use the same proposition with $D_{\text{IFS}}(A \cap X, A^c \cap X)$ and $D_{\text{IFS}}(A \cap X^c, A^c \cap X^c)$, we obtain that

$$\begin{aligned} D_{\text{IFS}}(A \cap X, A^c \cap X) &= D_{\text{IFS}}(A \cap X \cap Y, A^c \cap X \cap Y) \\ &\quad + D_{\text{IFS}}(A \cap X \cap Y^c, A^c \cap X \cap Y^c), \\ D_{\text{IFS}}(A \cap X^c, A^c \cap X^c) &= D_{\text{IFS}}(A \cap X^c \cap Y, A^c \cap X^c \cap Y) \\ &\quad + D_{\text{IFS}}(A \cap X^c \cap Y^c, A^c \cap X^c \cap Y^c). \end{aligned}$$

Hence,

$$\begin{aligned} D_{\text{IFS}}(A, A^c) &= D_{\text{IFS}}(A \cap X \cap Y, A^c \cap X \cap Y) \\ &\quad + D_{\text{IFS}}(A \cap X \cap Y^c, A^c \cap X \cap Y^c) \\ &\quad + D_{\text{IFS}}(A \cap X^c \cap Y, A^c \cap X^c \cap Y) \\ &\quad + D_{\text{IFS}}(A \cap X^c \cap Y^c, A^c \cap X^c \cap Y^c). \end{aligned}$$

Let us study each of the summands in the right-hand-side separately. For the first one, we have that

$$\begin{aligned} \mu_{A \cap X \cap Y}(\omega) &= \begin{cases} \mu_A(\omega) & \text{if } \mu_A(\omega) \leq 0.5 \leq \nu_A(\omega) \text{ and } \mu_B(\omega) \leq \nu_B(\omega), \\ 0 & \text{otherwise,} \end{cases} \\ \nu_{A \cap X \cap Y}(\omega) &= \begin{cases} \nu_A(\omega) & \text{if } \mu_A(\omega) \leq 0.5 \leq \nu_A(\omega) \text{ and } \mu_B(\omega) \leq \nu_B(\omega), \\ 1 & \text{otherwise.} \end{cases} \end{aligned}$$

However, if $\omega \in X \cap Y$, taking into account that $A \lll B$, it holds that

$$\mu_A(\omega) \leq \mu_B(\omega) \leq 0.5 \leq \nu_B(\omega) \leq \nu_A(\omega)$$

and therefore,

$$A \cap X \cap Y \subseteq B \subseteq B^c \cap X \cap Y \subseteq A^c \cap X \cap Y.$$

Now, applying Lemma 5.5 we obtain that

$$D_{\text{IFS}}(A \cap X \cap Y, A^c \cap X \cap Y) \geq D_{\text{IFS}}(B \cap X \cap Y, B^c \cap X \cap Y).$$

Let us consider next the second term

$$\begin{aligned} \mu_{A \cap X \cap Y^c}(\omega) &= \begin{cases} \mu_A(\omega) & \text{if } \mu_A(\omega) \leq 0.5 \leq \nu_A(\omega) \text{ and } \nu_B(\omega) \leq \mu_B(\omega), \\ 0 & \text{otherwise,} \end{cases} \\ \nu_{A \cap X \cap Y^c}(\omega) &= \begin{cases} \nu_A(\omega) & \text{if } \mu_A(\omega) \leq 0.5 \leq \nu_A(\omega) \text{ and } \nu_B(\omega) \leq \mu_B(\omega), \\ 1 & \text{otherwise,} \end{cases} \end{aligned}$$

However, if $\omega \in X \cap Y^c$, since $A \lll B$ it holds that

$$\mu_A(\omega) \leq \nu_B(\omega) \leq 0.5 \leq \mu_B(\omega) \leq \nu_A(\omega)$$

whence

$$A \cap X \cap Y^c \subseteq B^c \cap X \cap Y^c \subseteq B \cap X \cap Y^c \subseteq A^c \cap X \cap Y^c$$

and if we apply Lemma 5.5 we obtain that

$$D_{\text{IFS}}(A \cap X \cap Y^c, A^c \cap X \cap Y^c) \geq D_{\text{IFS}}(B \cap X \cap Y^c, B^c \cap X \cap Y^c).$$

Consider next the third summand

$$\begin{aligned} \mu_{A \cap X^c \cap Y}(\omega) &= \begin{cases} \mu_A(\omega) & \text{if } \nu_A(\omega) \leq 0.5 \leq \mu_A(\omega) \text{ and } \mu_B(\omega) \leq \nu_B(\omega), \\ 0 & \text{otherwise,} \end{cases} \\ \nu_{A \cap X^c \cap Y}(\omega) &= \begin{cases} \nu_A(\omega) & \text{if } \nu_A(\omega) \leq 0.5 \leq \mu_A(\omega) \text{ and } \mu_B(\omega) \leq \nu_B(\omega), \\ 1 & \text{otherwise.} \end{cases} \end{aligned}$$

If $\omega \in X^c \cap Y$, since $A \lll B$, it holds that

$$\nu_A(\omega) \leq \mu_B(\omega) \leq 0.5 \leq \nu_B(\omega) \leq \mu_A(\omega)$$

whence

$$A^c \cap X^c \cap Y \subseteq B \cap X^c \cap Y \subseteq B^c \cap X^c \cap Y \subseteq A \cap X^c \cap Y.$$

Applying Lemma 5.5, we obtain that

$$D_{\text{IFS}}(A \cap X^c \cap Y, A^c \cap X^c \cap Y) \geq D_{\text{IFS}}(B \cap X^c \cap Y, B^c \cap X^c \cap Y)$$

Finally, consider the fourth term:

$$\begin{aligned} \mu_{A \cap X^c \cap Y^c}(\omega) &= \begin{cases} \mu_A(\omega) & \text{if } \nu_A(\omega) \leq 0.5 \leq \mu_A(\omega) \text{ and } \nu_B(\omega) \leq \mu_B(\omega), \\ 0 & \text{otherwise,} \end{cases} \\ \nu_{A \cap X^c \cap Y^c}(\omega) &= \begin{cases} \nu_A(\omega) & \text{if } \nu_A(\omega) \leq 0.5 \leq \mu_A(\omega) \text{ and } \nu_B(\omega) \leq \mu_B(\omega), \\ 1 & \text{otherwise.} \end{cases} \end{aligned}$$

If $\omega \in X^c \cap Y^c$, taking into account that $A \lll B$, it holds that:

$$\nu_A(\omega) \leq \nu_B(\omega) \leq 0.5 \leq \mu_B(\omega) \leq \mu_A(\omega).$$

Then, using Lemma 5.5 we obtain that

$$D_{\text{IFS}}(A \cap X^c \cap Y^c, A^c \cap X^c \cap Y^c) \geq D_{\text{IFS}}(B \cap X^c \cap Y^c, B^c \cap X^c \cap Y^c)$$

and therefore

$$\begin{aligned} D_{\text{IFS}}(A, A^c) &= D_{\text{IFS}}(A \cap X \cap Y, A^c \cap X \cap Y) \\ &\quad + D_{\text{IFS}}(A \cap X \cap Y^c, A^c \cap X \cap Y^c) \\ &\quad + D_{\text{IFS}}(A \cap X^c \cap Y, A^c \cap X^c \cap Y) \\ &\quad + D_{\text{IFS}}(A \cap X^c \cap Y^c, A^c \cap X^c \cap Y^c) \\ &\geq D_{\text{IFS}}(B \cap X \cap Y, B^c \cap X \cap Y) \\ &\quad + D_{\text{IFS}}(B \cap X \cap Y^c, B^c \cap X \cap Y^c) \\ &\quad + D_{\text{IFS}}(B \cap X^c \cap Y, B^c \cap X^c \cap Y) \\ &\quad + D_{\text{IFS}}(B \cap X^c \cap Y^c, B^c \cap X^c \cap Y^c) = D_{\text{IFS}}(B, B^c). \end{aligned}$$

This completes the proof. ■

The above result implies that the lower the fuzziness, the greater the divergence between an IF-set and its complementary. Moreover, the divergence is maximum when the IF-set is crisp.

Proposition 5.35 *If V and Z are two crisp sets and D_{IFS} is a local IF-divergence,*

$$D_{\text{IFS}}(V, V^c) = D_{\text{IFS}}(Z, Z^c).$$

In addition, if $A, B \in \text{IFSS}(\Omega)$, then $D_{\text{IFS}}(A, B) \leq D_{\text{IFS}}(Z, Z^c)$.

Proof: Note that, by IF-loc.2 of Theorem 5.29 $h_{\text{IFS}}(1, 0, 0, 1) = h_{\text{IFS}}(0, 1, 1, 0)$, and therefore

$$D_{\text{IFS}}(V, V^c) = n \cdot h_{\text{IFS}}(1, 0, 0, 1) = D_{\text{IFS}}(Z, Z^c).$$

Now, taking into account that $h_{\text{IFS}}(1, 0, 0, 1) \geq h_{\text{IFS}}(x_1, x_2, y_1, y_2)$, since by IF-loc.3 and IF-loc.4:

$$\begin{aligned} h_{\text{IFS}}(1, 0, 0, 1) &\geq h_{\text{IFS}}(x_1, 0, 0, 1) \geq h_{\text{IFS}}(x_1, x_2, 0, 1) \\ &\geq h_{\text{IFS}}(x_1, x_2, 0, y_2) \geq h_{\text{IFS}}(x_1, x_2, y_1, y_2), \end{aligned}$$

we have that

$$\begin{aligned} D_{\text{IFS}}(A, B) &= \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)) \\ &\leq \sum_{i=1}^n h_{\text{IFS}}(1, 0, 0, 1) = D_{\text{IFS}}(Z, Z^c). \end{aligned} \quad \blacksquare$$

We have seen that every IF-divergence is also an IF-dissimilarity, and therefore it satisfies that $D_{\text{IFS}}(A, C) \geq \max(D_{\text{IFS}}(A, B), D_{\text{IFS}}(B, C))$ for every IF-sets A, B and C such that $A \subseteq B \subseteq C$. In the following proposition we obtain a similar result for local IF-divergences with less restrictive conditions.

Proposition 5.36 *Let D_{IFS} be a local IF-divergence. If for every $\omega \in \Omega$ either*

$$\mu_A(\omega) \leq \mu_B(\omega) \leq \mu_C(\omega) \text{ and } \nu_A(\omega) \geq \nu_B(\omega) \geq \nu_C(\omega),$$

or

$$\mu_A(\omega) \geq \mu_B(\omega) \geq \mu_C(\omega) \text{ and } \nu_A(\omega) \leq \nu_B(\omega) \leq \nu_C(\omega),$$

then $D_{\text{IFS}}(A, C) \geq \max(D_{\text{IFS}}(A, B), D_{\text{IFS}}(B, C))$.

Proof: Since the IF-divergence is local we can apply properties IF-loc.3 and IF-loc.4,

and we obtain the following:

$$\begin{aligned}
D_{\text{IFS}}(A, C) &= \sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_C(\omega_i), \nu_C(\omega_i)) \\
&\geq \max \left(\sum_{i=1}^n h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)), \right. \\
&\quad \left. \sum_{i=1}^n h_{\text{IFS}}(\mu_B(\omega_i), \nu_B(\omega_i), \mu_C(\omega_i), \nu_C(\omega_i)) \right) \\
&= \max(D_{\text{IFS}}(A, B), D_{\text{IFS}}(B, C)). \quad \blacksquare
\end{aligned}$$

In Proposition 5.27 we proved that, if D_{IFS} is an IF-divergence, then D_{IFS}^ϕ is also an IF-divergence, where $D_{\text{IFS}}^\phi(A, B) = \phi(D_{\text{IFS}}(A, B))$ and ϕ is a increasing function such that $\phi(0) = 0$. In particular, if D_{IFS} is a local IF-divergence, D_{IFS}^ϕ is local if and only if ϕ is linear. Next we derive a similar method to build local IF-divergences from local IF-divergences.

Proposition 5.37 *Let D_{IFS} be a local IF-divergence, and let $\phi : [0, \infty) \rightarrow [0, \infty)$ be a increasing function such that $\phi(0) = 0$. Then, the function $D_{\text{IFS}, \phi}$, defined by*

$$D_{\text{IFS}, \phi}(A, B) = \sum_{i=1}^n \phi(h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i))),$$

is a local IF-divergence.

Proof: Immediate using the properties of ϕ and taking into account that h_{IFS} satisfies the properties IF-loc.1 to IF-loc.5. \blacksquare

To conclude this section, we relate local IF-divergences and real distances.

Proposition 5.38 *Consider a distance $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ satisfying*

$$\max(d(x, y), d(y, z)) \leq d(x, z)$$

for $x < y < z$. Then, for every increasing function $\phi : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ such that $\phi(0, 0) = 0$, the function $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ defined by:

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n \phi(d(\mu_A(\omega_i), \mu_B(\omega_i)), d(\nu_A(\omega_i), \nu_B(\omega_i)))$$

is a local IF-divergence.

Proof: Using Theorem 5.42, it suffices to prove that the function

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = \phi(d(x_1, y_1), d(x_2, y_2))$$

satisfies the properties IF-loc.1 to IF-loc.5.

IF-loc.1: Consider $(x, y) \in \mathcal{T} = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}$. Since d is a distance, $d(x, x) = d(y, y) = 0$, and therefore

$$h_{\text{IFS}}(x, y, x, y) = \phi(d(x, x), d(y, y)) = \phi(0, 0) = 0.$$

IF-loc.2: Take (x_1, x_2) and (y_1, y_2) in \mathcal{T} . Since d is a distance, $d(x_1, y_1) = d(y_1, x_1)$ and $d(x_2, y_2) = d(y_2, x_2)$, whence

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = \phi(d(x_1, y_1), d(x_2, y_2)) = \phi(d(y_1, x_1), d(y_2, x_2)) = h_{\text{IFS}}(y_1, y_2, x_1, x_2).$$

IF-loc.3: Consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$ such that $x_1 \leq z \leq y_1$. Applying the hypothesis on d ,

$$d(x_1, y_1) \geq \max(d(x_1, z), d(z, y_1))$$

whence

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = \phi(d(x_1, y_1), d(x_2, y_2)) \geq \phi(d(x_1, z), d(x_2, y_2)) = h_{\text{IFS}}(x_1, x_2, z, y_2).$$

Moreover, if $(x_2, z), (y_2, z) \in \mathcal{T}$, then $\max(x_2, y_2) + z \leq 1$ and it holds that:

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = \phi(d(x_1, y_1), d(x_2, y_2)) \geq \phi(d(z, y_1), d(x_2, y_2)) = h_{\text{IFS}}(z, x_2, y_1, y_2).$$

IF-loc.4: Let $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$ such that $x_2 \leq z \leq y_2$. Applying the hypothesis on d ,

$$d(x_2, y_2) \geq \max(d(x_2, z), d(z, y_2)).$$

Since ϕ is increasing in each component:

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = \phi(d(x_1, y_1), d(x_2, y_2)) \geq \phi(d(x_1, y_1), d(x_2, z)) = h_{\text{IFS}}(x_1, x_2, y_1, z).$$

Moreover, if $(x_1, z), (y_1, z) \in \mathcal{T}$, it holds that $\max(x_1, y_1) + z \leq 1$ and then:

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = \phi(d(x_1, y_1), d(x_2, y_2)) \geq \phi(d(x_1, y_1), d(z, y_2)) = h_{\text{IFS}}(x_1, z, y_1, y_2).$$

IF-loc.5: Finally, consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$. Applying our hypothesis on d , it holds that:

$$d(z, z) = 0 \leq \min(d(x_1, y_1), d(x_2, y_2)).$$

Then, if $(x_2, z), (y_2, z) \in \mathcal{T}$, it holds that $\max(x_2, y_2) + z \leq 1$, and since ϕ is increasing in each component, it follows that

$$h_{\text{IFS}}(z, x_2, z, y_2) = \phi(d(z, z), d(x_2, y_2)) \leq \phi(d(x_1, y_1), d(x_2, y_2)) = h_{\text{IFS}}(x_1, x_2, y_1, y_2).$$

Moreover, if $(x_1, z), (x_2, z) \in \mathcal{T}$, then $\max(x_1, y_1) + z \leq 1$, and since ϕ is increasing in each component, it holds that:

$$h_{\text{IFS}}(x_1, z, y_1, z) = \phi(d(x_1, y_1), d(z, z)) \leq \phi(d(x_1, y_1), d(x_2, y_2)) = h_{\text{IFS}}(x_1, x_2, y_1, y_2).$$

Thus, h_{IFS} satisfies properties IF-loc.1 to IF-loc.5. Applying Theorem 5.29, we conclude that D_{IFS} is a local IF-divergence. ■

Let us see an example of an application of this result.

Example 5.39 Consider the distance d defined by $d(x, y) = |x - y|$, and the increasing function $\phi(x, y) = \frac{x+y}{2^n}$, that satisfies $\phi(0, 0) = 0$. Then, we can define the function $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ defined by

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n \phi(d(\mu_A(\omega_i), \mu_B(\omega_i)), d(\nu_A(\omega_i), \nu_B(\omega_i)))$$

for every $A, B \in \text{IFSS}(\Omega)$ is an IF-divergence. In fact, if we input the values of ϕ and d , D_{IFS} becomes

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|,$$

i.e., we obtain Hong and Kim IF-divergence D_C (see 5.1.3).◆

Examples of local IF-divergences

In this section we are going to study which of the examples of IF-divergences of Section 5.1.3 are in particular local IF-divergences.

Let us begin with the Hamming distance (see Section 5.1.3). It is defined by:

$$l_{\text{IFS}}(A, B) = \sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)| + |\pi_A(\omega_i) - \pi_B(\omega_i)|.$$

Consider two IF-sets A and B , and an element $\omega_i \in \Omega$. We have to see that the difference $l_{\text{IFS}}(A, B) - l_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\})$ only depends on $\mu_A(\omega_i), \mu_B(\omega_i), \nu_A(\omega_i)$ and $\nu_B(\omega_i)$. Note that, since $\mu_{A \cup \{\omega_i\}}(\omega_i) = \mu_{B \cup \{\omega_i\}}(\omega_i) = 1$ and $\nu_{A \cup \{\omega_i\}}(\omega_i) = \nu_{B \cup \{\omega_i\}}(\omega_i) = 0$, $l_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\})$ takes the following value:

$$l_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = \sum_{j \neq i} |\mu_A(\omega_j) - \mu_B(\omega_j)| + |\nu_A(\omega_j) - \nu_B(\omega_j)| + |\pi_A(\omega_j) - \pi_B(\omega_j)|$$

whence

$$l_{\text{IFS}}(A, B) - l_{\text{IFS}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)| + |\pi_A(\omega_i) - \pi_B(\omega_i)| = h_{\text{IFS}}(\mu_A(\omega_i), \nu_A(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)).$$

Thus, l_{IFS} is a local IF-divergence whose associated function h_{IFS} is given by:

$$h_{\text{IFS}}(x_1, x_2, y_1, y_2) = |x_1 - y_1| + |x_2 - y_2| + |x_1 + x_2 - y_1 - y_2|.$$

Moreover, the normalized Hamming distance, defined by $l_{\text{nIFS}}(A, B) = \frac{1}{n}l_{\text{IFS}}(A, B)$, is also a local IF-divergence. The reason is that $l_{\text{nIFS}}(A, B) = \phi(l_{\text{IFS}}(A, B))$, where $\phi(x) = \frac{x}{n}$, and we have already mentioned that in that case l_{nIFS} is local if and only if ϕ is linear.

Let us next study the Hausdorff distance for IF-sets (see Section 5.1.3), which is given by:

$$d_{\text{H}}(A, B) = \sum_{i=1}^n \max(|\mu_A(\omega_i) - \mu_B(\omega_i)|, |\nu_A(\omega_i), \nu_B(\omega_i)|).$$

Consider $\omega_i \in \Omega$, and let A and B be two IF-sets. As we have done in the previous case, $d_{\text{H}}(A \cup \{\omega_i\}, B \cup \{\omega_i\})$ is given by

$$d_{\text{H}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = \sum_{j \neq i} \max(|\mu_A(\omega_j) - \mu_B(\omega_j)|, |\nu_A(\omega_j), \nu_B(\omega_j)|),$$

taking into account that $A \cup \{\omega_i\}$ and $B \cup \{\omega_i\}$ are given by:

$$\begin{aligned} A \cup \{\omega_i\} &= \{(\omega_j, \mu_A(\omega_j), \nu_A(\omega_j)), (\omega_i, 1, 0) \mid j \neq i\}. \\ B \cup \{\omega_i\} &= \{(\omega_j, \mu_B(\omega_j), \nu_B(\omega_j)), (\omega_i, 1, 0) \mid j \neq i\}. \end{aligned}$$

Hence, $d_{\text{H}}(A, B) - d_{\text{H}}(A \cup \{\omega_i\}, B \cup \{\omega_i\})$ is given by

$$d_{\text{H}}(A, B) - d_{\text{H}}(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = \max(|\mu_A(\omega_i) - \mu_B(\omega_i)|, |\nu_A(\omega_i) - \nu_B(\omega_i)|).$$

Therefore, the Hamming distance for IF-sets is a local IF-divergence, whose associated function h_{dH} is given by

$$h_{\text{dH}}(x_1, x_2, y_1, y_2) = \max(|x_1 - y_1|, |x_2 - y_2|).$$

The same applies to the normalized Hausdorff distance, since it is a linear transformation of the Hausdorff distance.

Consider now the IF-divergences defined by Hong and Kim, D_{C} and D_{L} (see Section 5.1.3), given by

$$\begin{aligned} D_{\text{C}}(A, B) &= \frac{1}{2n} \sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|. \\ D_{\text{L}}(A, B) &= \frac{1}{4n} \sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i) - \nu_A(\omega_i) + \nu_B(\omega_i)| \\ &\quad + |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|. \end{aligned}$$

Let us see that both IF-divergences are local. Consider two IF-sets A and B and an element $\omega_i \in \Omega$, and let us compute $D_C(A \cup \{\omega_i\}, B \cup \{\omega_i\})$ and $D_L(A \cup \{\omega_i\}, B \cup \{\omega_i\})$.

$$D_C(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = \frac{1}{2n} \sum_{j \neq i} |\mu_A(\omega_j) - \mu_B(\omega_j)| + |\nu_A(\omega_j) - \nu_B(\omega_j)|.$$

$$D_L(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = \frac{1}{4n} \sum_{j \neq i} |\mu_A(\omega_j) - \mu_B(\omega_j) - \nu_A(\omega_j) + \nu_B(\omega_j)|$$

$$+ |\mu_A(\omega_j) - \mu_B(\omega_j)| + |\nu_A(\omega_j) - \nu_B(\omega_j)|.$$

Then,

$$D_C(A, B) - D_C(A \cup \{\omega_i\}, B \cup \{\omega_i\}) = |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|.$$

$$D_L(A, B) - D_L(A \cup \{\omega_i\}, B \cup \{\omega_i\})$$

$$= |\mu_A(\omega_j) - \mu_B(\omega_j) - \nu_A(\omega_j) + \nu_B(\omega_j)|$$

$$+ |\mu_A(\omega_j) - \mu_B(\omega_j)| + |\nu_A(\omega_j) - \nu_B(\omega_j)|.$$

Thus, both IF-divergences are local, and their respective functions h_{D_C} and h_{D_L} are:

$$h_{D_C}(x_1, x_2, y_1, y_2) = |x_1 - y_1| + |x_2 - y_2|.$$

$$h_{D_L}(x_1, x_2, y_1, y_2) = |x_1 - y_1 - x_2 + y_2| + |x_1 - y_1| + |x_2 - y_2|.$$

In summary, Hamming and Hausdorff distances and the IF-divergences of Hong and Kim are local IF-divergences. It can be checked that the other examples of IF-divergences are not local.

5.1.5 IF-divergences Vs Divergences

Some of the studies presented until now in this chapter are inspired in the concept of fuzzy divergence introduced by Montes et al. ([160]).

Definition 5.40 ([160]) *Let Ω be an universe. A map $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ is a divergence if it satisfies the following conditions:*

- Div.1:* $D(A, A) = 0$ for every $A \in FS(\Omega)$.
- Div.2:* $D(A, B) = D(B, A)$ for every $A, B \in FS(\Omega)$.
- Div.3:* $D(A \cap C, B \cap C) \leq D(A, B)$, for every $A, B, C \in FS(\Omega)$.
- Div.4:* $D(A \cup C, B \cup C) \leq D(A, B)$, for every $A, B, C \in FS(\Omega)$.

Montes et al ([160]) also investigated the local property for fuzzy divergences.

Definition 5.41 ([160, Def. 3.2]) *A divergence measure defined on a finite universe is a local divergence, or it is said to fulfill the local property, if for every $A, B \in FS(\Omega)$ and every $\omega \in \Omega$ we have that:*

$$D(A, B) - D(A \cup \{\omega\}, B \cup \{\omega\}) = h(A(\omega), B(\omega)),$$

Local fuzzy divergences were characterized as follows.

Theorem 5.42 ([160, Prop. 3.4]) *A map $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ defined on a finite universe $\Omega = \{\omega_1, \dots, \omega_n\}$ is a local divergence if and only if there is a function $h : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ such that*

$$D(A, B) = \sum_{i=1}^n h(A(\omega_i), B(\omega_i)),$$

and

loc.1: $h(x, y) = h(y, x)$, for every $(x, y) \in [0, 1]^2$.

loc.2: $h(x, x) = 0$ for every $x \in [0, 1]$.

loc.3: $h(x, z) \geq \max(h(x, y), h(y, z))$, for every $x, y, z \in [0, 1]$ such that $x < y < z$.

In this section we are going to study the relationship between divergences and IF-divergences. We shall provide some methods to derive IF-divergences from divergences and vice versa. Moreover, we shall investigate under which conditions the property of being local is preserved under these transformations.

From IF-divergences to fuzzy divergences

Consider an IF-divergence $D_{\text{IFS}} : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$ defined on a finite universe $\Omega = \{\omega_1, \dots, \omega_n\}$. Recall that every fuzzy set A is in particular an IF-set, whose membership and non-membership functions are $\mu_A(\omega_i) = A(\omega_i)$ and $\nu_A(\omega_i) = 1 - A(\omega_i)$, respectively. Hence, if A and B are two fuzzy sets, we can compute its divergence D as:

$$D(A, B) = D_{\text{IFS}}(A, B).$$

Proposition 5.43 *If D_{IFS} is an IF-divergence, the map $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ given by*

$$D(A, B) = D_{\text{IFS}}(A, B)$$

is a divergence for fuzzy sets. Moreover, if D_{IFS} satisfies axiom IF-Div.5, then D satisfies axiom Div.5, and if D_{IFS} is local, then so is D .

Proof: Let us prove that D is a divergence, i.e., that it satisfies axioms Diss.1 to Div.4.

Diss.1: Let A be a fuzzy set. Then:

$$D(A, A) = D_{\text{IFS}}(A, A) = 0.$$

Div.2: Let A and B be two fuzzy sets. Since they are in particular IF-sets, $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A)$, and therefore:

$$D(A, B) = D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A) = D(B, A).$$

Div.3: Let A , B and C be fuzzy sets. Again, since they are in particular IF-sets, it holds that $D_{\text{IFS}}(A \cap C, B \cap C) \leq D_{\text{IFS}}(A, B)$. Then:

$$D(A \cap C, B \cap C) = D_{\text{IFS}}(A \cap C, B \cap C) \leq D_{\text{IFS}}(A, B) = D(A, B).$$

Div.4: Similarly to Div.3, consider fuzzy sets A , B and C . Since they are in particular IF-sets, they satisfy $D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B)$, whence

$$D(A \cup C, B \cup C) = D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B) = D(A, B).$$

Thus, D is a divergence for fuzzy sets. Assume now that D_{IFS} satisfies IF-Div.5, i.e.,

$$D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c) \text{ for every } A, B \in \text{IFSs}(\Omega).$$

Then, in particular, D satisfies axiom Div.5

$$D(A, B) = D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c) = D(A^c, B^c),$$

for every $A, B \in \text{FS}(\Omega)$. Assume now that D_{IFS} is a local IF-divergence. Then:

$$\begin{aligned} D(A, B) - D(A \cup \{\omega\}, B \cup \{\omega\}) &= D_{\text{IFS}}(A, B) - D_{\text{IFS}}(A \cup \{\omega\}, B \cup \{\omega\}) \\ &= \tilde{h}(A(\omega), 1 - A(\omega), B(\omega), 1 - B(\omega)) = h(A(\omega), B(\omega)), \end{aligned}$$

where $h(x, y) = \tilde{h}(x, 1 - x, y, 1 - y)$. Consequently, D is a local divergence between fuzzy sets. \blacksquare

Remark 5.44 *The function D defined in the previous proposition is in fact a composition of some functions:*

$$D : \text{FS}(\Omega) \times \text{FS}(\Omega) \xrightarrow{i} \text{IFSs}(\Omega) \times \text{IFSs}(\Omega) \xrightarrow{D_{\text{IFS}}} \mathbb{R}$$

where $i(A, B)$ stands for the inclusion of $\text{FS}(\Omega) \times \text{FS}(\Omega)$ on $\text{IFSs}(\Omega) \times \text{IFSs}(\Omega)$. \blacklozenge

Remark 5.45 *If we look at the proof of Proposition 5.43, we see that, in order to prove that D satisfies axiom Div. i , for $i \in \{1, 2\}$ it is enough for D_{IFS} to satisfy axiom IF-Diss. i . Moreover, if D_{IFS} satisfies axiom IF-Div. j , for $j \in \{3, 4\}$, then D also satisfies axiom Div. j . In fact, if for instance D_{IFS} is not an IF-divergence, but it satisfies IF-Diss.1, IF-Diss.2 and IF-Div.3, we cannot assure that D is a divergence. However, we know that D satisfies axioms Div.1, IF-Div.2 and IF-Div.3. \blacklozenge*

The above method of deriving divergences from IF-divergences seems to be natural. Let us show how it can be used in a few examples.

Example 5.46 Consider the Hamming distance for IF-sets that we have already studied in Section 5.1.3, given by:

$$l_{\text{IFS}}(A, B) = \frac{1}{2} \sum_{i=1}^n (|\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)| + |\pi_A(\omega_i) - \pi_B(\omega_i)|).$$

If we consider A and B two fuzzy sets, the divergence D defined in the previous proposition is:

$$D_1(A, B) = \sum_{i=1}^n |A(\omega_i) - B(\omega_i)|.$$

Recall that the function:

$$l_{\text{FS}}(A, B) = \sum_{i=1}^n |A(\omega_i) - B(\omega_i)|, \quad \forall A, B \in \text{FS}(\Omega)$$

is known as the Hamming distance for fuzzy sets. Then, from the Hamming distance for IF-sets we obtain the Hamming distance for fuzzy sets. Moreover, if we consider the normalized Hamming distance for IF-sets, we also obtain the normalized Hamming distance, defined by $l_{\text{nIFS}}(A, B) = \frac{1}{n} l_{\text{IFS}}$, for fuzzy sets.

Consider now the Hausdorff distance (see Section 5.1.3) for IF-sets:

$$d_{\text{H}}(A, B) = \sum_{i=1}^n \max(|\mu_A(\omega_i) - \mu_B(\omega_i)|, |\nu_A(\omega_i) - \nu_B(\omega_i)|).$$

Given two fuzzy sets A and B , if we apply Proposition 5.43 we obtain the Hamming distance for fuzzy sets:

$$\begin{aligned} D_2(A, B) &= d_{\text{H}}(A, B) = \sum_{i=1}^n \max(|A(\omega_i) - B(\omega_i)|, |(1 - A(\omega_i)) - (1 - B(\omega_i))|) \\ &= \sum_{i=1}^n |A(\omega_i) - B(\omega_i)| = l_{\text{FS}}(A, B). \end{aligned}$$

Moreover, if we consider the normalized Hausdorff distance, we obtain the normalized Hamming distance:

$$\begin{aligned} D_3(A, B) &= d_{\text{nH}}(A, B) = \frac{1}{n} \sum_{i=1}^n \max(|A(\omega_i) - B(\omega_i)|, |(1 - A(\omega_i)) - (1 - B(\omega_i))|) \\ &= \frac{1}{n} \sum_{i=1}^n |A(\omega_i) - B(\omega_i)| = l_{\text{nFS}}(A, B). \end{aligned}$$

Thus, both the Hamming distance and the Hausdorff distance for IF-sets produce the same divergence for fuzzy sets: the Hamming distance for fuzzy sets.

However, if we consider the IF-divergences of Hong and Kim (see Section 5.1.3), defined by:

$$D_C(A, B) = \frac{1}{2n} \sum_{i=1}^n (|\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|);$$

$$D_L(A, B) = \frac{1}{4n} \sum_{i=1}^n (|(\mu_A(\omega_i) - \mu_B(\omega_i)) - (\nu_A(\omega_i) - \nu_B(\omega_i))| + |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|);$$

and we apply Proposition 5.43 we obtain also the normalized Hamming distance:

$$D_4(A, B) = D_C(A, B) = \frac{1}{2n} \sum_{i=1}^n (|A(\omega_i) - B(\omega_i)| + |(1 - A(\omega_i)) - (1 - B(\omega_i))|)$$

$$= \frac{1}{n} \sum_{i=1}^n |A(\omega_i) - B(\omega_i)| = l_{\text{nFS}}(A, B).$$

$$D_5(A, B) = D_L(A, B) = \frac{1}{4n} \sum_{i=1}^n (|(A(\omega_i) - B(\omega_i)) - (1 - A(\omega_i) - 1 + B(\omega_i))| + |A(\omega_i) - B(\omega_i)| + |1 - A(\omega_i) - 1 + B(\omega_i)|)$$

$$= \frac{1}{n} \sum_{i=1}^n |A(\omega_i) - B(\omega_i)| = l_{\text{nFS}}(A, B).$$

Thus, both Hamming and Hausdorff distances for IF-sets produce the Hamming distance for fuzzy sets, and the normalized Hamming and Hausdorff distances, and Hong and Kim dissimilarities for IF-sets produce the normalized Hamming distance for fuzzy sets. Consequently, all these IF-divergences can be seen as generalizations of the Hamming distance for fuzzy sets to the comparison of IF-sets. ♦

Example 5.47 Let us now consider the IF-divergence defined by Li et al. (see page 283 of Section 5.1.3):

$$D_O(A, B) = \frac{1}{\sqrt{2n}} \left(\sum_{i=1}^n (\mu_A(\omega_i) - \mu_B(\omega_i))^2 + (\nu_A(\omega_i) - \nu_B(\omega_i))^2 \right)^{\frac{1}{2}}.$$

If we use Proposition 5.43 in order to build a divergence for fuzzy sets from D_O , we obtain the normalized Euclidean distance for fuzzy sets:

$$D(A, B) = D_O(A, B) = \frac{1}{\sqrt{2n}} \left(\sum_{i=1}^n (A(\omega_i) - B(\omega_i))^2 + (1 - A(\omega_i) - 1 + B(\omega_i))^2 \right)^{\frac{1}{2}}$$

$$= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n (A(\omega_i) - B(\omega_i))^2 \right)^{\frac{1}{2}} = \sqrt{2} q_{\text{nFS}}(A, B).$$

Thus, both the normalized Euclidean distance for IF-sets and Li et al. IF-divergence are generalizations of the normalized Euclidean distance for fuzzy sets. Note however that the normalized Euclidean distance is not an IF-divergence (see Section 5.1.3), even though Li et al.'s dissimilarity is. ♦

Example 5.48 Consider now the IF-divergence defined by Mitchell (see Section 5.1.3):

$$D_{\text{HB}}(A, B) = \frac{1}{2\sqrt[p]{n}} \left(\left(\sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |\nu_A(\omega_i) - \nu_B(\omega_i)|^p \right)^{\frac{1}{p}} \right).$$

Applying Proposition 5.43, we obtain the following divergence for fuzzy sets:

$$\begin{aligned} D_1(A, B) = D_{\text{HB}}(A, B) &= \frac{1}{2\sqrt[p]{n}} \left(\left(\sum_{i=1}^n |A(\omega_i) - B(\omega_i)|^p \right)^{\frac{1}{p}} \right. \\ &\quad \left. + \left(\sum_{i=1}^n |(1 - A(\omega_i)) - (1 - B(\omega_i))|^p \right)^{\frac{1}{p}} \right) = \frac{1}{\sqrt[p]{n}} \left(\sum_{i=1}^n |A(\omega_i) - B(\omega_i)|^p \right)^{\frac{1}{p}}. \end{aligned}$$

If we now consider the IF-Divergence D_e^p of Liang and Shi (see Section 5.1.3), defined by:

$$D_e^p(A, B) = \frac{1}{2\sqrt[p]{n}} \left(\sum_{i=1}^n (|\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|)^p \right)^{\frac{1}{p}}$$

and apply Proposition 5.43, we obtain the following divergence:

$$\begin{aligned} D_2(A, B) &= D_e^p(A, B) \\ &= \frac{1}{2\sqrt[p]{n}} \left(\sum_{i=1}^n (|A(\omega_i) - B(\omega_i)| + |(1 - A(\omega_i)) - (1 - B(\omega_i))|)^p \right)^{\frac{1}{p}} \\ &= \frac{1}{\sqrt[p]{n}} \left(\sum_{i=1}^n |A(\omega_i) - B(\omega_i)|^p \right)^{\frac{1}{p}}. \end{aligned}$$

Note that $D_1(A, B) = D_2(A, B)$. Thus, both D_{HB} and D_e^p produce the same divergence between fuzzy sets, and therefore both of them can be seen as a generalization of the divergence D_1 . ♦

Although the method proposed in Proposition 5.43 seems to be very natural, there is another possible, albeit less intuitive, way of deriving divergences from IF-divergences, that we detail next.

Proposition 5.49 The function $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ defined by

$$D(A, B) = D_{\text{IFS}}(\tilde{A}, \tilde{B}),$$

where D_{IFS} is an IF-divergence, is a divergence for fuzzy sets, where \tilde{A} and \tilde{B} are given by:

$$\begin{aligned}\tilde{A} &= \{(\omega, A(\omega), 0) \mid \omega \in \Omega\} \in \text{IFSS}(\Omega). \\ \tilde{B} &= \{(\omega, B(\omega), 0) \mid \omega \in \Omega\} \in \text{IFSS}(\Omega).\end{aligned}$$

However, although D_{IFS} satisfies IF-Div.5, D may not satisfy Div.5.

Proof: Let us see that D satisfies the divergence axioms.

Diss.1: Let A be a fuzzy set. Then $\tilde{A} = \{(\omega, A(\omega), 0) \mid \omega \in \Omega\}$, and therefore, as D_{IFS} is an IF-divergence,

$$D(A, A) = D_{\text{IFS}}(\tilde{A}, \tilde{A}) = 0.$$

Diss.2: Let A and B be two fuzzy sets. Then

$$D(A, B) = D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A) = D(B, A),$$

because D_{IFS} is symmetric.

Div.3: Consider $A, B, C \in \text{IFSS}(\Omega)$. Since D_{IFS} is an IF-divergence, $D_{\text{IFS}}(\tilde{A} \cap \tilde{C}, \tilde{B} \cap \tilde{C}) \leq D_{\text{IFS}}(\tilde{A}, \tilde{B})$. Moreover,

$$\begin{aligned}\widetilde{A \cap C} &= \{(\omega, \min(\mu_A(\omega), \mu_C(\omega)), 0) \mid \omega \in \Omega\} = \tilde{A} \cap \tilde{C}. \\ \widetilde{B \cap C} &= \{(\omega, \min(\mu_B(\omega), \mu_C(\omega)), 0) \mid \omega \in \Omega\} = \tilde{B} \cap \tilde{C},\end{aligned}$$

whence

$$\begin{aligned}D(A \cap C, B \cap C) &= D_{\text{IFS}}(\widetilde{A \cap C}, \widetilde{B \cap C}) \\ &= D_{\text{IFS}}(\tilde{A} \cap \tilde{C}, \tilde{B} \cap \tilde{C}) \leq D_{\text{IFS}}(\tilde{A}, \tilde{B}) = D(A, B).\end{aligned}$$

Div.4: The proof is similar to the previous one. Consider three fuzzy sets A, B and C . We know that $D_{\text{IFS}}(\tilde{A} \cup \tilde{C}, \tilde{B} \cup \tilde{C}) \leq D_{\text{IFS}}(\tilde{A}, \tilde{B})$. Moreover,

$$\begin{aligned}\widetilde{A \cup C} &= \{(\omega, \max(\mu_A(\omega), \mu_C(\omega)), 0) \mid \omega \in \Omega\} = \tilde{A} \cup \tilde{C}. \\ \widetilde{B \cup C} &= \{(\omega, \max(\mu_B(\omega), \mu_C(\omega)), 0) \mid \omega \in \Omega\} = \tilde{B} \cup \tilde{C}.\end{aligned}$$

Then, axiom Div.4 is satisfied, because:

$$\begin{aligned}D(A \cup C, B \cup C) &= D_{\text{IFS}}(\widetilde{A \cup C}, \widetilde{B \cup C}) \\ &= D_{\text{IFS}}(\tilde{A} \cup \tilde{C}, \tilde{B} \cup \tilde{C}) \leq D_{\text{IFS}}(\tilde{A}, \tilde{B}) = D(A, B).\end{aligned}$$

Hence, D is a divergence for fuzzy sets. Assume now that D_{IFS} satisfies axiom IF-Div.5 and let us show that in that case D may not satisfy Div.5. Consider a singleton universe $\Omega = \{\omega\}$, and the function $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ defined by

$$\begin{aligned}D_{\text{IFS}}(A, B) &= |\max(\mu_A(\omega) - 0.5, 0) - \max(\mu_B(\omega) - 0.5, 0)| \\ &\quad + |\max(\nu_A(\omega) - 0.5, 0) - \max(\nu_B(\omega) - 0.5, 0)|.\end{aligned}$$

Let us see that D_{IFS} is an IF-divergence.

IF-Diss.1: Let A be an IF-set. Trivially

$$\begin{aligned} |\max(\mu_A(\omega) - 0.5, 0) - \max(\mu_A(\omega) - 0.5, 0)| &= 0 \text{ and} \\ |\max(\nu_A(\omega) - 0.5, 0) - \max(\nu_A(\omega) - 0.5, 0)| &= 0, \end{aligned}$$

and therefore $D_{\text{IFS}}(A, A) = 0$.

IF-Diss.2: Let A and B be two IF-sets. Then it follows from the definition that $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(B, A)$.

IF-Div.3: Let A , B and C be three IF-sets. We must prove the following inequality:

$$\begin{aligned} &|\max(\mu_A(\omega) - 0.5, 0) - \max(\mu_B(\omega) - 0.5, 0)| + \\ &|\max(\nu_A(\omega) - 0.5, 0) - \max(\nu_B(\omega) - 0.5, 0)| \geq \\ &|\max(\mu_{A \cap C}(\omega) - 0.5, 0) - \max(\mu_{B \cap C}(\omega) - 0.5, 0)| + \\ &|\max(\nu_{A \cap C}(\omega) - 0.5, 0) - \max(\nu_{B \cap C}(\omega) - 0.5, 0)|. \end{aligned}$$

This follows from Lemma A.5 in Appendix A.

IF-Div.4: Similarly, if A , B and C are three IF-sets, condition IF-Div.4 holds if and only if:

$$\begin{aligned} &|\max(\mu_A(\omega) - 0.5, 0) - \max(\mu_B(\omega) - 0.5, 0)| + \\ &|\max(\nu_A(\omega) - 0.5, 0) - \max(\nu_B(\omega) - 0.5, 0)| \geq \\ &|\max(\mu_{A \cup C}(\omega) - 0.5, 0) - \max(\mu_{B \cup C}(\omega) - 0.5, 0)| + \\ &|\max(\nu_{A \cup C}(\omega) - 0.5, 0) - \max(\nu_{B \cup C}(\omega) - 0.5, 0)|, \end{aligned}$$

and this follows from Lemma A.5 in Appendix A.

Hence, D_{IFS} is an IF-divergence. Moreover, it also trivially satisfies axiom IF-Div.5.

Consider the divergence derived in this proposition:

$$D(A, B) = D_{\text{IFS}}(\{(\omega, A(\omega), 0)\}, \{(\omega, B(\omega), 0)\}) = |\max(A(\omega) - 0.5, 0) - \max(B(\omega) - 0.5, 0)|.$$

Although D_{IFS} satisfies IF-Div.5, D does not fulfill Div.5: if we consider the fuzzy sets A and B given by

$$\begin{aligned} A = \{(\omega, 0.3)\} &\Rightarrow A^c = \{(\omega, 0.7)\}, \text{ and} \\ B = \{(\omega, 0.4)\} &\Rightarrow B^c = \{(\omega, 0.6)\} \end{aligned}$$

then it holds that $D(A, B) = 0 \neq 0.1 = D(A^c, B^c)$. ■

Although this second method for deriving divergences from IF-divergences is also valid, for us the first one seems to be more natural; besides, we have shown that some of the most important examples of divergences can be obtained applying this method to the corresponding IF-divergences.

From fuzzy divergences to IF-divergences

Consider now a divergence $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ between fuzzy sets define on a finite space $\Omega = \{\omega_1, \dots, \omega_n\}$, and let us study how to derive an IF-divergence from it. Consider two IF-sets A and B . Each of them can be decomposed into two fuzzy sets as follows:

$$\begin{aligned} A &= \{(\omega_i, \mu_A(\omega_i), \nu_A(\omega_i) \mid i = 1, \dots, n\} \in IFSs(\Omega) \\ &\Rightarrow \begin{cases} A_1 = \{(\omega_i, \mu_A(\omega_i) \mid i = 1, \dots, n\} \in FS(\Omega) \subseteq IFSs(\Omega). \\ A_2 = \{(\omega_i, \nu_A(\omega_i)) \mid i = 1, \dots, n\} \in FS(\Omega) \subseteq IFSs(\Omega). \end{cases} \\ B &= \{(\omega_i, \mu_B(\omega_i), \nu_B(\omega_i) \mid i = 1, \dots, n\} \in IFSs(\Omega) \\ &\Rightarrow \begin{cases} B_1 = \{(\omega_i, \mu_B(\omega_i) \mid i = 1, \dots, n\} \in FS(\Omega) \subseteq IFSs(\Omega). \\ B_2 = \{(\omega_i, \nu_B(\omega_i)) \mid i = 1, \dots, n\} \in FS(\Omega) \subseteq IFSs(\Omega). \end{cases} \end{aligned}$$

Using the divergence D we can measure the divergence between the pairs of fuzzy sets (A_1, B_1) and (A_2, B_2) . In other words, we have the divergence between the membership degrees and the non-membership degrees; in order to compute the divergence between A and B it only remains to combine these two divergences.

Theorem 5.50 *Let D be a divergence for fuzzy sets, and let $f : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ be a mapping satisfying the following two properties:*

$$\begin{aligned} f1: & f(0, 0) = 0; \\ f2: & f(\cdot, t) \text{ and } f(t, \cdot) \text{ are increasing for every } t \in [0, \infty); \end{aligned}$$

then, the function $D_{IFS} : IFSs(\Omega) \times IFSs(\Omega) \rightarrow \mathbb{R}$ defined by

$$D_{IFS}(A, B) = f(D(A_1, B_1), D(A_2, B_2)), \text{ for every } A, B \in IFSs(\Omega),$$

is an IF-divergence. Moreover, if D is a local divergence, then D_{IFS} is also a local IF-divergence if f has the form: $f(x, y) = \alpha x + \beta y$, for some $\alpha, \beta \geq 0$.

Finally, if f is symmetric then D_{IFS} fulfills axiom IF-Div.5 (regardless of whether D satisfies or not axiom Div.5), and if f is not symmetric, then although D satisfies Div.5, D_{IFS} may not satisfy IF-Div.5.

Proof: We begin by showing that D_{IFS} is an IF-divergence.

IF-Diss.1: Let A be an IF-set. Applying the definition of D_{IFS} we obtain that:

$$D_{IFS}(A, A) = f(D(A_1, A_1), D(A_2, A_2)) = f(0, 0) \stackrel{f1}{=} 0.$$

IF-Diss.2: Let A, B be IF-sets, and let us prove that $D_{IFS}(A, B) = D_{IFS}(B, A)$.

$$\begin{aligned} D_{IFS}(A, B) &= f(D(A_1, B_1), D(A_2, B_2)) \\ &= f(D(B_1, A_1), D(B_2, A_2)) = D_{IFS}(B, A). \end{aligned}$$

IF-Div.3: Consider the IF-sets A , B and C , and let us prove that $D_{\text{IFS}}(A \cap C, B \cap C) \leq D_{\text{IFS}}(A, B)$. Let us note the following:

$$\begin{aligned} A \cap C &= \{(\omega, \mu_{A \cap C}(\omega), \nu_{A \cap C}(\omega)) \mid \omega \in \Omega\} \\ &= \{(\omega, \min(\mu_A(\omega), \mu_C(\omega)), \max(\nu_A(\omega), \nu_C(\omega))) \mid \omega \in \Omega\} \\ &\Rightarrow \begin{cases} (A \cap C)_1 = \{(\omega, \min(\mu_A(\omega), \mu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega). \\ (A \cap C)_2 = \{(\omega, \max(\nu_A(\omega), \nu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega). \end{cases} \end{aligned}$$

Similarly, we also obtain that

$$\begin{aligned} (B \cap C)_1 &= \{(\omega, \min(\mu_B(\omega), \mu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega). \\ (B \cap C)_2 &= \{(\omega, \max(\nu_B(\omega), \nu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega). \end{aligned}$$

Since D is a divergence for fuzzy sets, applying Div.3 we obtain that:

$$D(A \cap C_1, B \cap C_1) = D((A \cap C)_1, (B \cap C)_1) \leq D(A_1, B_1),$$

where $C_1 = \mu_C$, and applying Div.4,

$$D(A \cup C_2, B \cup C_2) = D((A \cap C)_2, (B \cap C)_2) \leq D(A_2, B_2),$$

where $C_2 = \nu_C$. From these properties, $D_{\text{IFS}}(A \cap C, B \cap C) \leq D_{\text{IFS}}(A, B)$

$$\begin{aligned} D_{\text{IFS}}(A \cap C, B \cap C) &= f(D((A \cap C)_1, (B \cap C)_1), D((A \cup C)_2, (B \cup C)_2)) \\ &\leq f(D(A_1, B_1), D(A_2, B_2)) = D_{\text{IFS}}(A, B). \end{aligned}$$

IF-Div.4: Let us prove that $D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B)$ for every IF-sets A , B and C , similarly to the previous point. We have that

$$\begin{aligned} A \cup C_1 &= \{(\omega, \max(\mu_A(\omega), \mu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega), \\ A \cup C_2 &= \{(\omega, \min(\nu_A(\omega), \nu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega), \\ B \cup C_1 &= \{(\omega, \max(\mu_B(\omega), \mu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega), \\ B \cup C_2 &= \{(\omega, \min(\nu_B(\omega), \nu_C(\omega))) \mid \omega \in \Omega\} \in FS(\Omega). \end{aligned}$$

Applying Div.4,

$$D(A \cup C_1, B \cup C_1) \leq D(A_1, B_1),$$

and Div.3 implies that:

$$D(A \cup C_2, B \cup C_2) \leq D(A_2, B_2).$$

Using these two inequalities, we can prove that $D_{\text{IFS}}(A \cup C, B \cup C) \leq D_{\text{IFS}}(A, B)$

$$\begin{aligned} D_{\text{IFS}}(A \cup C, B \cup C) &= f(D(A \cup C_1, B \cup C_1), D(A \cup C_2, B \cup C_2)) \\ &\leq f(D(A_1, B_1), D(A_2, B_2)) = D_{\text{IFS}}(A, B). \end{aligned}$$

Hence, D_{IFS} is an IF-divergence. Assume now that f is symmetric, i.e., $f(x, y) = f(y, x)$ for every $(x, y) \in [0, 1]^2$, then it is immediate that D_{IFS} satisfies axiom IF-Div.5, that is, $D_{\text{IFS}}(A, B) = D_{\text{IFS}}(A^c, B^c)$ for every $A, B \in \text{IFSS}(\Omega)$, since:

$$D_{\text{IFS}}(A, B) = f(D(A_1, B_1), D(A_2, B_2)) = f(D(A_2, B_2), D(A_1, B_1)) = D_{\text{IFS}}(A^c, B^c).$$

However, assume that f is not symmetric, and let us give an example of divergence D that fulfills axiom Div.5, such that D_{IFS} does not satisfies IF-Div.5. Consider the normalized Hamming divergence for fuzzy sets:

$$l_{\text{FS}}(A, B) = \frac{1}{n} \sum_{i=1}^n |A(\omega_i) - B(\omega_i)|,$$

and let f be given by: $f(x, y) = \alpha x + \beta y$, where $\alpha \neq \beta$, for example $\alpha = 1$ and $\beta = 0$. Then:

$$D_{\text{IFS}}(A, B) = \frac{1}{n} \sum_{i=1}^n (\alpha |\mu_A(\omega_i) - \mu_B(\omega_i)| + \beta |\nu_A(\omega_i) - \nu_B(\omega_i)|)$$

is an IF-divergence. Obviously D satisfies axiom Div.5, but D_{IFS} does not satisfy IF-Div.5; to see this, it suffices to consider the IF-sets

$$A = \{(\omega, 0.6, 0.2) \mid \omega \in \Omega\} \text{ and } B = \{(\omega, 0.5, 0.4) \mid \omega \in \Omega\}.$$

Then it holds that

$$D_{\text{IFS}}(A, B) = \frac{1}{n} \sum_{i=1}^n (\alpha \cdot 0.1 + \beta \cdot 0.2) = \alpha \cdot 0.1 + \beta \cdot 0.2 = 0.1.$$

$$D_{\text{IFS}}(A^c, B^c) = \frac{1}{n} \sum_{i=1}^n (\alpha \cdot 0.2 + \beta \cdot 0.1) = \alpha \cdot 0.2 + \beta \cdot 0.1 = 0.2.$$

and therefore $D_{\text{IFS}}(A, B) \neq D_{\text{IFS}}(A^c, B^c)$.

Assume now that D is a local divergence, i.e., that there is a function h , such that

- loc.1: $h(x, y) = h(y, x)$, for every $(x, y) \in [0, 1]^2$;
- loc.2: $h(x, x) = 0$ for every $x \in [0, 1]$;
- loc.3: $h(x, z) \geq \max(h(x, y), h(y, z))$, for every $x, y, z \in [0, 1]$ such that $x < y < z$;

for which D can be expressed by:

$$D(A, B) = \sum_{i=1}^n h(A(\omega_i), B(\omega_i)).$$

Then, D_{IFS} is given by

$$D_{\text{IFS}}(A, B) = f \left(\sum_{i=1}^n h(\mu_A(\omega_i), \mu_B(\omega_i)), \sum_{i=1}^n h(\nu_A(\omega_i), \nu_B(\omega_i)) \right).$$

Let us see that if f is linear then D_{IFS} is a local IF-divergence. In such a case, D_{IFS} has the following form:

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n \alpha \cdot h(\mu_A(\omega_i), \mu_B(\omega_i)) + \beta \cdot h(\nu_A(\omega_i), \nu_B(\omega_i)),$$

and if we define \tilde{h} by:

$$\tilde{h}(x_1, y_1, x_2, y_2) = \alpha \cdot h(x_1, x_2) + \beta \cdot h(y_1, y_2)$$

then it suffices to show that \tilde{h} satisfies properties (i)-(iv) to deduce that D_{IFS} is a local IF-divergence. Let us see that this is indeed the case:

IF-loc.1: Consider $(x, y) \in [0, 1]^2$. By hypothesis it holds that $h(x, x) = h(y, y) = 0$, and then:

$$\tilde{h}(x, y, x, y) = \alpha h(x, x) + \beta h(y, y) = 0.$$

IF-loc.2: Consider (x_1, x_2) and (y_1, y_2) in \mathcal{T} . Then $h(x_1, y_1) = h(y_1, x_1)$ and $h(x_2, y_2) = h(y_2, x_2)$, whence

$$\begin{aligned} \tilde{h}(x_1, x_2, y_1, y_2) &= \alpha h(x_1, y_1) + \beta h(x_2, y_2) \\ &= \alpha h(y_1, x_1) + \beta h(y_2, x_2) = \tilde{h}(y_1, y_2, x_1, x_2). \end{aligned}$$

IF-loc.3: Take now $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$ such that $x_1 \leq z \leq y_1$. Then, loc.3 implies that:

$$h(x_1, y_1) \geq \max(h(x_1, z), h(z, y_1)),$$

whence

$$\begin{aligned} \tilde{h}(x_1, x_2, y_1, y_2) &= \alpha \cdot h(x_1, y_1) + \beta \cdot h(x_2, y_2) \geq \alpha \cdot \max(h(x_1, z), h(z, y_1)) + \beta \cdot h(x_2, y_2) \\ &= \max(\tilde{h}(x_1, x_2, z, y_2), \tilde{h}(z, x_2, y_1, y_2)). \end{aligned}$$

In particular, $\tilde{h}(x_1, x_2, y_1, y_2) \geq \tilde{h}(x_1, x_2, z, y_2)$ and, if $(x_2, z), (y_2, z) \in \mathcal{T}$, then $\max(x_2 + z, y_2 + z) \leq 1$ and $\tilde{h}(x_1, x_2, y_1, y_2) \geq \tilde{h}(z, x_2, y_1, y_2)$.

IF-loc.4: Consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$ such that $x_2 \leq z \leq y_2$. Applying property loc.3 we see that

$$h(x_2, y_2) \geq \max(h(x_2, z), h(z, y_2))$$

and therefore

$$\begin{aligned} \tilde{h}(x_1, x_2, y_1, y_2) &= \alpha h(x_1, y_1) + \beta h(x_2, y_2) \geq \alpha h(x_1, y_1) + \beta \max(h(x_2, z), h(z, y_2)) \\ &= \max(\tilde{h}(x_1, x_2, y_1, z), \tilde{h}(x_1, z, y_1, y_2)). \end{aligned}$$

IF-loc.5: Consider $(x_1, x_2), (y_1, y_2) \in \mathcal{T}$ and $z \in [0, 1]$. By loc.1, we know that $h(z, z) = 0$. Then:

$$\begin{aligned}\tilde{h}(z, x_2, z, y_2) &= \alpha h(z, z) + \beta h(x_2, y_2) = \beta h(x_2, y_2) \\ &\leq \alpha h(x_1, y_1) + \beta h(x_2, y_2) = \tilde{h}(x_1, x_2, y_1, y_2). \\ \tilde{h}(x_1, z, y_1, z) &= \alpha h(x_1, y_1) + \beta h(z, z) = \alpha h(x_1, y_1) \\ &\leq \alpha h(x_1, y_1) + \beta h(x_2, y_2) = \tilde{h}(x_1, x_2, y_1, y_2).\end{aligned}$$

Thus, D_{IFS} is a local divergence. ■

Remark 5.51 In a similar way, it is possible to prove that, if D_1 and D_2 are two divergences for fuzzy sets, and if $f : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ is an increasing function with $f(0, 0) = 0$, then the function $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ defined by:

$$D_{\text{IFS}}(A, B) = f(D_1(\mu_A, \mu_B), D_2(\nu_A, \nu_B))$$

for every $A, B \in \text{IFSS}(\Omega)$, is an IF-divergence. ◆

If in particular we consider the function $f(x, y) = x$ we obtain the following result.

Corollary 5.52 Let D be a map $D : \text{FS}(\Omega) \times \text{FS}(\Omega) \rightarrow \mathbb{R}$, and consider the function $f : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ given by $f(x, y) = x$. Define $D_{\text{IFS}} : \text{IFSS}(\Omega) \times \text{IFSS}(\Omega) \rightarrow \mathbb{R}$ by:

$$D_{\text{IFS}}(A, B) = f(D(A_1, B_1), D(A_2, B_2)), \text{ for every } A, B \in \text{IFSS}(\Omega).$$

Then, if D satisfies axiom Div. i ($i \in \{1, 2\}$), then D_{IFS} satisfies axiom IF-Diss. i , and if D satisfies axiom Div. j ($j \in \{3, 4\}$), D_{IFS} satisfies axiom IF-v. j . In particular, if D is a divergence for fuzzy sets, then D_{IFS} is an IF-divergence. Moreover, if D is local, then D_{IFS} is also a local IF-divergence. However, D_{IFS} may not satisfy the property IF-Div.5 even if D satisfies Div.5.

Proof:

- Let us assume that D satisfies Diss.1. Then, D_{IFS} satisfies IF-Diss.1 since:

$$D_{\text{IFS}}(A, A) = D(A_1, A_1) = 0.$$

- Let us assume that D satisfies Diss.2. Then, D_{IFS} is also symmetric since:

$$D_{\text{IFS}}(A, B) = D(A_1, B_1) = D(B_1, A_1) = D_{\text{IFS}}(B, A).$$

- Let us assume that D satisfies Div.3, and let us see that $D_{\text{IFS}}(A \cap C, B \cap C) \leq D_{\text{IFS}}(A, B)$ for every IF-sets A, B and C .

$$D_{\text{IFS}}(A \cap C, B \cap C) = D(A \cap C_1, B \cap C_1) \leq D(A_1, B_1) = D_{\text{IFS}}(A, B).$$

- Finally, assume that D satisfies Div.4. Then also D_{IFS} satisfies axiom IF-Div.4, since for every A , B and C it holds that:

$$D_{\text{IFS}}(A \cup C, B \cup C) = D(A \cup C_1, B \cup C_1) \leq D(A_1, B_1) = D_{\text{IFS}}(A, B).$$

Thus, if D is a divergence for fuzzy sets, then D_{IFS} is also an IF-divergence. Moreover, taking into account the previous theorem and that f is a linear function, if D is a local divergence, then D_{IFS} is also a local IF-divergence. Furthermore, we have seen in that result that a sufficient condition for D_{IFS} to satisfy IF-Div.5 is that f is symmetric, which is not the case for $f(x, y) = x$. Then, we cannot assure D_{IFS} to satisfy IF-Div.5. ■

Using the previous results we can give some examples of IF-divergences.

Example 5.53 Consider the function $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ defined by:

$$D(A, B) = \sum_{\omega \in \Omega} h(A(\omega), B(\omega)),$$

where $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by:

$$h(x, y) = \begin{cases} 0 & \text{if } x = y. \\ 1 - xy & \text{if } x \neq y. \end{cases}$$

Montes proved that this function satisfies Div.1, Div.2 and Div.3 (see [159]). Then, if we apply Theorem 5.50 with the function $f(x, y) = x$, we conclude that the function D_1 satisfies IF-Div.1, IF-Div.2 and IF-Div.3.

Similarly, we can consider the function

$$h(x, y) = \begin{cases} 0 & \text{if } x = y, \\ xy & \text{if } x \neq y, \end{cases}$$

and $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ defined by:

$$D(A, B) = \sum_{\omega \in \Omega} h(A(\omega), B(\omega)).$$

Montes et al. ([159]) proved that D satisfies Div.1, Div.2 and Div.4. Then, applying Theorem 5.50 with the function $f(x, y) = x$, we conclude that the function D_2 they generate satisfies IF-Diss.1, IF-Diss.2 and IF-Div.4.

These two functions D_1 and D_2 were used in Example 5.22, and there we have proved that they are not IF-divergences. ♦

Example 5.54 In Equation (5.2), we considered a function $D : FS(\Omega) \times FS(\Omega) \rightarrow \mathbb{R}$ defined on the space $\Omega = \{\omega\}$ by:

$$D(A, B) = D_{IFS}(A, B) = |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)|^2.$$

The Hamming distance for fuzzy sets, l_{FS} , is known to be a divergence for fuzzy sets. Then, applying Theorem 5.50 to this divergence and the function $f(x, y) = x + y^2$, we obtain the function of Equation (5.2), and therefore we conclude that it is an IF-divergence. ♦

Assume now that we have an IF-divergence D_{IFS} . Using Theorem 5.50 we can build a divergence D for fuzzy sets. On the other hand, Proposition 5.43 allows us to derive another IF-divergence D_{IFS}^* . We next investigate under which conditions these two IF-divergences coincide.

Remark 5.55 Let us consider D_{IFS} an IF-divergence. Let D be the divergence determined by Proposition 5.43:

$$D(A, B) = D_{IFS}(A, B), \text{ for every } A, B \in FS(\Omega).$$

and let D_{IFS}^* be the IF-divergence derived from D by means of Theorem 5.50:

$$D_{IFS}^*(A, B) = f(D(A_1, B_1), D(A_2, B_2)), \text{ for every } A, B \in IFSs(\Omega).$$

Then, $D_{IFS} = D_{IFS}^*$ if and only if for every $A, B \in IFSs(\Omega)$ it holds that:

$$D_{IFS}(A, B) = f(D_{IFS}(A_1, B_1), D_{IFS}(A_2, B_2)).$$

Similarly, let D be a divergence for fuzzy sets. Using Theorem 5.50 we can build an IF-divergence D_{IFS} , and applying Proposition 5.43, from D_{IFS} we can derive a divergence D^* . Again, we want to determine if we recover our initial divergence.

Theorem 5.56 Let D be a divergence for fuzzy sets, and let D_{IFS} be the IF-divergence derived from D by means of Theorem 5.50, given by

$$D_{IFS}(A, B) = f(D(A_1, B_1), D(A_2, B_2)) \quad \forall A, B \in IFSs(\Omega).$$

Let D^* be the divergence derived from D_{IFS} by means of Proposition 5.43:

$$D^*(A, B) = D_{IFS}(A, B), \text{ for every } A, B \in FS(\Omega).$$

Then, $D = D^*$ if and only if $f(x, y) = x$ for every $(x, y) \in [0, 1]^2$.

Proof: Let us compute the expression of D^* :

$$D^*(A, B) = D_{\text{IFS}}(A, B) = f(D(A, B), D(A^c, B^c)),$$

for every $A, B \in FS(\Omega)$. Thus, $D(A, B) = D^*(A, B)$ for every $A, B \in FS(\Omega)$ if and only if:

$$D(A, B) = f(D(A, B), D(A^c, B^c)),$$

and this is equivalent to $f(x, y) = x$ for every $(x, y) \in [0, 1]^2$. ■

Let us see how Remark 5.55 and Theorem 5.56 apply to the Hamming distance for fuzzy sets and the IF-divergence of Hong and Kim.

Example 5.57 *Let us consider the Hamming distance for fuzzy sets:*

$$l_{\text{FS}}(A, B) = \sum_{i=1}^n |A(\omega_i) - B(\omega_i)|, \text{ for every } A, B \in FS(\Omega).$$

Applying Theorem 5.50, we can build an IF-divergence from l_{FS} :

$$D_{\text{IFS}}(A, B) = f\left(\sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)|, \sum_{i=1}^n |\nu_A(\omega_i) - \nu_B(\omega_i)|\right),$$

and using Proposition 5.43, we can derive from D_{IFS} another divergence D for fuzzy sets:

$$D(A, B) = f\left(\sum_{i=1}^n |A(\omega_i) - B(\omega_i)|, |A(\omega_i) - B(\omega_i)|\right).$$

Then, $D(A, B) = l_{\text{FS}}(A, B)$ if and only if $f(x, x) = x$. In particular D and l_{FS} are the same divergence if $f(x, y) = \frac{x+y}{2}$.

Consider now the IF-divergence D_C defined by Hong and Kim in Section 5.1.3:

$$D_C(A, B) = \frac{1}{2} \sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)| + |\nu_A(\omega_i) - \nu_B(\omega_i)|.$$

Using Proposition 5.43 we can build a divergence for fuzzy sets:

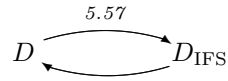
$$D(A, B) = D_{\text{IFS}}(A, B) = \sum_{i=1}^n |A(\omega_i) - B(\omega_i)| = l_{\text{FS}}(A, B).$$

If we now apply Theorem 5.50, we can build other IF-divergence given by:

$$\begin{aligned} D_{\text{IFS}}(A, B) &= f(D(A_1, B_1), D(A_2, B_2)) \\ &= f\left(\sum_{i=1}^n |\mu_A(\omega_i) - \mu_B(\omega_i)|, \sum_{i=1}^n |\nu_A(\omega_i) - \nu_B(\omega_i)|\right). \end{aligned}$$

Thus, we conclude that $D_{\text{IFS}}(A, B) = D_C(A, B)$ if and only if $f(x, y) = \frac{x+y}{2}$. ◆

Corollary 5.58 *Let D be a divergence for fuzzy sets. Then, the diagram:*



commutes if and only if $f(x, y) = x$ and

$$D_{\text{IFS}}(A, B) = D(A_1, B_1), \text{ for every } A, B \in \text{IFSS}(\Omega).$$

Proof: On the one hand, from Theorem 5.56 we know that $f(x, y) = x$. Moreover, from Remark 5.55 the following equation must hold:

$$\begin{aligned}
 D_{\text{IFS}}(A, B) &= f(D_{\text{IFS}}(A_1, B_1), D_{\text{IFS}}(A_2, B_2)) = D_{\text{IFS}}(A_1, B_1) \\
 &= f(D(A_1, B_1), D(A_2, B_2)) = D(A_1, B_1).
 \end{aligned}$$

Thus, for every $A, B \in \text{IFSS}(\Omega)$ it must hold that:

$$D_{\text{IFS}}(A, B) = D(A_1, B_1). \blacksquare$$

5.2 Connecting IVF-sets and imprecise probabilities

This section is devoted to investigate the relationship between IF-sets and Imprecise Probabilities. In fuzzy set theory, it is well known ([217]) that there exists a connection between fuzzy sets and possibility measures. In fact, given a normalized fuzzy set μ_A , it defines a possibility distribution with associated possibility measure Π defined by:

$$\Pi(B) = \sup_{x \in B} \mu_A(x).$$

Conversely, given a possibility measure Π with associated possibility distribution π , it defines a fuzzy set with membership function μ .

In this section, we shall assume first of all that the IVF-sets are defined on a probability space. Thus, any IVF-set defines a random set, and then the probabilistic information of the IVF-set can be summarized by means of the set of distributions of the measurable selections. In this framework, we investigate in which situations the probabilistic information can be equivalently represented by the set of probabilities that dominate the lower probability induced by the random interval, and the conditions under which the upper probability induced by the random interval is a possibility measure.

Afterwards, we shall investigate other possible relationships between IVF-sets and imprecise probabilities. For instance, we shall see that the definition of probability for IVF-set given by Grzegorzewski and Mrowka ([86]) becomes a particular case in our theory. We also investigate how a one-to-one relation could be defined between IVF-sets, p-boxes and clouds.

5.2.1 Probabilistic information of IVF-sets

In this section we shall assume that IVF-sets are defined on a probability space. Then, they define random sets. We investigate how the probabilistic information of a IVF-set can be summarized by means of Imprecise Probabilities.

Since formally IVF-sets and IF-sets are equivalent, as we saw in Section 2.3, we shall denote IVF-sets by:

$$\{[\mu_A(\omega), 1 - \nu_A(\omega)] : \omega \in \Omega\},$$

where μ_A and ν_A refer the membership and non-membership degree of the associated IF-set.

IVFS as random intervals

As we mentioned in Section 2.3, an IVF-set can be regarded as a model for the imprecise knowledge about the membership function of a fuzzy set A , in the sense that for every ω in the possibility space Ω , its membership degree belongs to the interval $[\mu_A(\omega), 1 - \nu_A(\omega)]$. Hence, we can equivalently represent the IVF-set I_A by means of a multi-valued mapping $\Gamma_A : \Omega \rightarrow \mathcal{P}([0, 1])$, where

$$\Gamma_A(\omega) := [\mu_A(\omega), 1 - \nu_A(\omega)]. \quad (5.5)$$

If the intuitionistic fuzzy set is defined on a probability space (Ω, \mathcal{A}, P) , then the probabilistic information encoded by the multi-valued mapping Γ_A can be summarized by means of its lower and upper probabilities $P_{*\Gamma_A}, P_{\Gamma_A}^*$. Recall that, from Equation (2.22), for any subset B in the Borel σ -field $\beta_{[0,1]}$, its lower and upper probabilities are given by

$$P_{*\Gamma_A}(B) := P(\{\omega : \Gamma_A(\omega) \subseteq B\})$$

and

$$P_{\Gamma_A}^*(B) := P(\{\omega : \Gamma_A(\omega) \cap B \neq \emptyset\}).$$

We need to make two clarifications here: the first one is that the images of the multi-valued mapping Γ_A are non-empty, as a consequence of the restriction $\mu_A \leq 1 - \nu_A$ in the definition of IVF-sets; the second is that, in order to be able to define the lower and upper probabilities $P_{*\Gamma_A}, P_{\Gamma_A}^*$, the multi-valued mapping Γ_A needs to be *strongly measurable* ([88]), which in this case ([129]) means that the mappings

$$\mu_A, \nu_A : \Omega \rightarrow [0, 1]$$

must be $\mathcal{A} - \beta_{[0,1]}$ -measurable.

If we assume that the ‘true’ membership function imprecisely specified by means of the IVF-set is $\mathcal{A} - \beta_{[0,1]}$ -measurable, then it must belong to the set of measurable selections of Γ_A (see Equation (2.21)):

$$S(\Gamma_A) := \{\phi : \Omega \rightarrow [0, 1] \text{ measurable} : \phi(\omega) \in [\mu_A(\omega), 1 - \nu_A(\omega)] \forall \omega \in \Omega\},$$

and as a consequence the probability measure it induces will belong to the set

$$P(\Gamma_A) := \{P_\phi : \phi \in S(\Gamma_A)\}.$$

Any probability measure in $P(\Gamma_A)$ is bounded by the upper probability $P_{\Gamma_A}^*$, and as a consequence the set $P(\Gamma_A)$ is included in the set $M(P_{\Gamma_A}^*)$ of probability measures that are dominated by $P_{\Gamma_A}^*$. As we have seen in Section 2.2.4, both sets are not equivalent in general; however, Proposition 2.45 shows several situations in which they coincide. Taking this result into account, we can establish the following conditions for the equality between the credal sets generated by an IVF-set.

Corollary 5.59 *Consider the initial space $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and $\Gamma_A : [0, 1] \rightarrow \mathcal{P}([0, 1])$ defined as in Equation (5.5). Then, the equality $M(P_{\Gamma_A}^*) = P(\Gamma_A)$ holds under any of the following conditions:*

- (a) *The membership function μ_A is increasing and the non-membership function ν_A is decreasing.*
- (b) *$\mu_A(\omega) = 0$ for any $\omega \in \Omega$.*
- (c) *For any $\omega, \omega' \in \Omega$, either $\Gamma_A(\omega) \leq \Gamma_A(\omega')$ or $\Gamma_A(\omega) \geq \Gamma_A(\omega')$, where $[a_1, b_1] \leq [a_2, b_2]$ if $a_1 \leq a_2$ and $b_1 \leq b_2$.*

The previous conditions can be interpreted as follows:

- (a) The greater the value of ω , the more evidence supports that ω belongs to A .
- (b) There is no evidence supporting that the elements belong to set A .
- (c) The intervals associated with the elements are ordered. In particular, this holds when the hesitation is constant.

Proof:

- (a) Condition (3a) of Proposition 2.45 assures that $M(P_{\Gamma_A}^*)$ and $P(\Gamma_A)$ coincide whenever the bounds of the random interval are increasing. In the particular case of IVF-set, this means that both μ_A and $1 - \nu_A$ are increasing, or equivalently, that μ_A is increasing and ν_A is decreasing.

(b) Condition (3b) of Proposition 2.45 assures that $M(P_{\Gamma_A}^*)$ and $P(\Gamma_A)$ coincide if the lower bound of the interval equals 0. In the case of IVF-sets, this means that $\mu_A = 0$.

(c) Condition (3c) of Proposition 2.45 assures that $M(P_{\Gamma_A}^*)$ and $P(\Gamma_A)$ coincide if the bounds of the interval are strictly comonotone. In the case of IVF-sets, the bounds of the interval, μ_A and $1 - \nu_A$, are comonotone if and only if $\Gamma_A(\omega) \geq \Gamma_A(\omega')$ or $\Gamma_A(\omega) \leq \Gamma_A(\omega')$ for any ω, ω' : assume that μ_A and $1 - \nu_A$ are comonotone, then $\mu_A(\omega) \geq \mu_A(\omega')$ if and only if $1 - \nu_A(\omega) \geq 1 - \nu_A(\omega')$ for every ω, ω' . Thus:

– If $\mu_A(\omega) > \mu_A(\omega')$, then $1 - \nu_A(\omega) > 1 - \nu_A(\omega')$, so

$$\Gamma_A(\omega) = [\mu_A(\omega), 1 - \nu_A(\omega)] > [\mu_A(\omega'), 1 - \nu_A(\omega')] = \Gamma_A(\omega').$$

– If $\mu_A(\omega) < \mu_A(\omega')$, then $1 - \nu_A(\omega) < 1 - \nu_A(\omega')$, so

$$\Gamma_A(\omega) = [\mu_A(\omega), 1 - \nu_A(\omega)] < [\mu_A(\omega'), 1 - \nu_A(\omega')] = \Gamma_A(\omega').$$

On the other hand, assume that either $\Gamma_A(\omega) \geq \Gamma_A(\omega')$ or $\Gamma_A(\omega) \leq \Gamma_A(\omega')$ for any ω, ω' . Then:

$$\begin{aligned} \Gamma_A(\omega) \geq \Gamma_A(\omega') &\Rightarrow \mu_A(\omega) \geq \mu_A(\omega') \text{ and } 1 - \nu_A(\omega) \geq 1 - \nu_A(\omega') \\ \Gamma_A(\omega) \leq \Gamma_A(\omega') &\Rightarrow \mu_A(\omega) \leq \mu_A(\omega') \text{ and } 1 - \nu_A(\omega) \leq 1 - \nu_A(\omega') \end{aligned}$$

and from this we deduce that μ_A and $1 - \nu_A$ are comonotone. ■

On the other hand, [129, Example 3.3] shows that the equality $P(\Gamma) = M(P_{\Gamma}^*)$ does not necessarily hold for all the random closed intervals, even when the initial probability space is non-atomic: it suffices to consider $(\Omega, \mathcal{A}, P) = ([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and $\Gamma : [0, 1] \rightarrow \mathcal{P}(\mathbb{R})$ given by

$$\Gamma(\omega) = [-\omega, \omega] \quad \forall \omega \in [0, 1].$$

It is easy to adapt the example to our context and deduce that there are intuitionistic fuzzy sets where the information about the membership function is not completely determined by the upper probability $P_{\Gamma_A}^*$: it would suffice to take $\Gamma_A : [0, 1] \rightarrow \mathcal{P}([0, 1])$ given by

$$\Gamma_A(\omega) = \left[0.5 - \frac{\omega}{2}, 0.5 + \frac{\omega}{2}\right] \quad \forall \omega \in [0, 1], \quad (5.6)$$

that is, to consider the IVF-set such that the membership and non-membership functions of its associated IF-set coincide and take the value $\frac{1-\omega}{2}$.

We have seen in Proposition 2.47 that the upper probability associated with a random set is a possibility measure if and only if the images of Γ are nested except for a null subset. In the particular case of the random closed intervals associated with an IVF-set, we deduce the following:

Corollary 5.60 *Let $\Gamma_A : \Omega \rightarrow \mathcal{P}([0, 1])$ be the random set defined in the probability space (Ω, \mathcal{A}, P) by Equation (5.5). Then, P_Γ^* is possibility measure if and only if there exists some $N \subseteq \Omega$ null such that μ_A and ν_A are comonotone on $\Omega \setminus N$.*

Proof: Assume that Γ_A is a possibility measure. Then, by Proposition 2.47, there is a null set N such that $\Gamma_A(\omega_1) \subseteq \Gamma_A(\omega_2)$ or $\Gamma_A(\omega_2) \subseteq \Gamma_A(\omega_1)$ for any $\omega_1, \omega_2 \in \Omega \setminus N$. Consider $\omega_1, \omega_2 \in \Omega \setminus N$, it holds that:

$$\begin{aligned} \Gamma_A(\omega_1) \subseteq \Gamma_A(\omega_2) &\Rightarrow [\mu_A(\omega_1), 1 - \nu_A(\omega_1)] \subseteq [\mu_A(\omega_2), 1 - \nu_A(\omega_2)] \\ &\Rightarrow \mu_A(\omega_1) \geq \mu_A(\omega_2) \text{ and } 1 - \nu_A(\omega_1) \leq 1 - \nu_A(\omega_2) \\ &\Rightarrow \mu_A(\omega_1) \geq \mu_A(\omega_2) \text{ and } \nu_A(\omega_1) \geq \nu_A(\omega_2) \\ \Gamma_A(\omega_2) \subseteq \Gamma_A(\omega_1) &\Rightarrow [\mu_A(\omega_2), 1 - \nu_A(\omega_2)] \subseteq [\mu_A(\omega_1), 1 - \nu_A(\omega_1)] \\ &\Rightarrow \mu_A(\omega_2) \geq \mu_A(\omega_1) \text{ and } 1 - \nu_A(\omega_2) \leq 1 - \nu_A(\omega_1) \\ &\Rightarrow \mu_A(\omega_2) \geq \mu_A(\omega_1) \text{ and } \nu_A(\omega_2) \geq \nu_A(\omega_1). \end{aligned}$$

Then, μ_A and ν_A are comonotone on $\Omega \setminus N$.

Conversely, assume that μ_A and ν_A are comonotone on $\Omega \setminus N$.

$$\begin{aligned} \text{If } \mu_A(\omega_1) \leq \mu_A(\omega_2) &\Rightarrow \nu_A(\omega_1) \leq \nu_A(\omega_2) \\ &\Rightarrow [\mu_A(\omega_1), 1 - \nu_A(\omega_1)] \subseteq [\mu_A(\omega_2), 1 - \nu_A(\omega_2)] \Rightarrow \Gamma_A(\omega_2) \subseteq \Gamma_A(\omega_1). \\ \text{If } \mu_A(\omega_2) \leq \mu_A(\omega_1) &\Rightarrow \nu_A(\omega_2) \leq \nu_A(\omega_1) \\ &\Rightarrow [\mu_A(\omega_2), 1 - \nu_A(\omega_2)] \subseteq [\mu_A(\omega_1), 1 - \nu_A(\omega_1)] \Rightarrow \Gamma_A(\omega_1) \subseteq \Gamma_A(\omega_2). \quad \blacksquare \end{aligned}$$

P-box induced by a IVF-set

The lower and upper probabilities $P_{*\Gamma_A}, P_{*\Gamma_A}^*$ summarize the probabilistic information about the probability distribution of the membership function of the IVF-set A . In particular we want to summarise the information about the distribution function of this variable, we must use the lower and upper distribution functions:

$$\underline{F}_A, \overline{F}_A : \Omega \rightarrow [0, 1],$$

where

$$\underline{F}_A(x) := P_{*\Gamma_A}([0, x]) = P(\{\omega : 1 - \nu_A(\omega) \leq x\}) = P_{\nu_A}([1 - x, 1]) \quad (5.7)$$

and

$$\overline{F}_A(x) := P_{*\Gamma_A}^*([0, x]) = P(\{\omega : \mu_A(\omega) \leq x\}) = P_{\mu_A}([0, x]). \quad (5.8)$$

When Ω is an ordered space (for instance if $\Omega = [0, 1]$), the lower and upper distribution functions $\underline{F}_A, \overline{F}_A$ can be used to determine a p -box. In that case, we shall refer to $(\underline{F}_A, \overline{F}_A)$ as the p -box on Ω associated with the intuitionistic fuzzy set A .

The lower and upper distribution functions also determine a set of probability measures:

$$M(\underline{F}_A, \overline{F}_A) := \{Q : \beta_{[0,1]} \rightarrow [0, 1] : \underline{F}_A(x) \leq F_Q(x) \leq \overline{F}_A(x) \forall x \in [0, 1]\},$$

where F_Q is the distribution function associated with the probability measure Q . It is immediate to see that the set $M(\underline{F}_A, \overline{F}_A)$ includes $M(P_{\Gamma_A}^*)$. However, the two sets do not coincide in general, and as a consequence the use of the lower and upper distribution functions may produce a loss of information, as we can see in the following example.

Example 5.61 Consider the random set of Equation (5.6), defined on $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ by $\Gamma_A(\omega) = [0.5 - \frac{\omega}{2}, 0.5 + \frac{\omega}{2}]$. Using Equation (2.23), we already know that the credal set $M(P_{\Gamma_A}^*)$ is given by:

$$M(P_{\Gamma_A}^*) = \{P \text{ probability} \mid P_{*\Gamma_A}(B) \leq P(B) \leq P_{\Gamma_A}^*(B) \text{ for any } B\}.$$

Let us now compute the form of the set $M(\underline{F}_A, \overline{F}_A)$:

$$\begin{aligned} \underline{F}_A(x) &= P_{*\Gamma_A}([0, x]) = P(\{\omega \in [0, 1] : \Gamma(\omega) \subseteq [0, x]\}) \\ &= P(\{\omega \in [0, 1] : \Gamma(\omega) = [0.5 - \frac{\omega}{2}, 0.5 + \frac{\omega}{2}] \subseteq [0, x]\}) \\ &= P(\{\omega \in [0, 1] : \omega \in [-1, 2x - 1]\}) \\ &= P(\{\omega \in [0, 1] : \omega \in [0, 2x - 1]\}) \\ &= \begin{cases} 0 & \text{if } x \leq \frac{1}{2}. \\ 2x - 1 & \text{otherwise.} \end{cases} \\ \overline{F}_A(x) &= P_{\Gamma_A}^*([0, x]) = P(\{\omega \in [0, 1] : \Gamma(\omega) \cap [0, x] \neq \emptyset\}) \\ &= P(\{\omega \in [0, 1] : [0.5 - \frac{\omega}{2}, 0.5 + \frac{\omega}{2}] \cap [0, x] \neq \emptyset\}) \\ &= \begin{cases} 2x & \text{if } x \leq \frac{1}{2}. \\ 1 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, the set $M(\underline{F}_A, \overline{F}_A)$ is formed by the probabilities whose associated cumulative distribution function is bounded by \underline{F}_A and \overline{F}_A .

Consider now the probability distribution associated with the cumulative distribution function F defined by:

$$F(x) = \begin{cases} \overline{F}_A(x) & \text{if } x \leq \frac{1}{4}. \\ \frac{1}{2} & \text{if } x \in (\frac{1}{4}, \frac{3}{4}]. \\ \underline{F}_A(x) & \text{if } x > \frac{3}{4}. \end{cases}$$

Its associated probability, P_F , belongs to $M(\underline{F}_A, \overline{F}_A)$. Now, let us check that P_F does not belong to $M(P_{\Gamma_A}^*)$. For this aim, note that:

$$\begin{aligned} P_{*\Gamma_A}([\frac{1}{4}, \frac{3}{4}]) &= P(\{\omega \in [0, 1] : \Gamma(\omega) \subseteq [\frac{1}{4}, \frac{3}{4}]\}) \\ &= P(\{\omega \in [0, 1] : [0.5 - \frac{\omega}{2}, 0.5 + \frac{\omega}{2}] \subseteq [\frac{1}{4}, \frac{3}{4}]\}) \\ &= P(\{\omega \in [0, 1] : \omega \in [0, \frac{1}{2}]\}) = \frac{1}{2}. \end{aligned}$$

This means that every probability P in $M(P_{\Gamma_A}^*)$ must hold that $P\left(\left[\frac{1}{4}, \frac{3}{4}\right]\right) \geq \frac{1}{2}$. However, $P_F\left(\left[\frac{1}{4}, \frac{3}{4}\right]\right) = 0$, and consequently $P_F \notin M(P_{\Gamma_A}^*)$.

We conclude that $M(\underline{F}_A, \overline{F}_A) \supset M(P_{\Gamma_A}^*)$. ♦

Nevertheless, there are non-trivial situations in which both sets coincide.

Example 5.62 Consider the initial space $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random set Γ_A defined from the IF-set I_A by:

$$\Gamma_A(\omega) = \begin{cases} \{\omega\} & \text{if } \omega \in [0, \frac{1}{4}] \cup (\frac{3}{4}, 1] \\ \left[\frac{1}{4}, \frac{3}{4}\right] & \text{otherwise.} \end{cases}$$

Thus, the membership and non-membership functions are given by:

$$\mu_A(\omega) = \begin{cases} \omega & \text{if } \omega \in [0, \frac{1}{4}] \cup (\frac{3}{4}, 1] \\ \frac{1}{4} & \text{otherwise,} \end{cases}$$

and

$$\nu_A(\omega) = \begin{cases} 1 - \omega & \text{if } \omega \in [0, \frac{1}{4}] \cup (\frac{3}{4}, 1] \\ \frac{1}{4} & \text{otherwise.} \end{cases}$$

Then, the lower and upper cdfs \underline{F}_A and \overline{F}_A are given by:

$$\underline{F}_A(x) = \begin{cases} x & \text{if } x \in [0, \frac{1}{4}), \\ \frac{1}{4} & \text{if } x \in [\frac{1}{4}, \frac{3}{4}), \\ x & \text{if } x \in [\frac{3}{4}, 1], \end{cases} \quad \text{and} \quad \overline{F}_A(x) = \begin{cases} x & \text{if } x \in [0, \frac{1}{4}). \\ \frac{3}{4} & \text{if } x \in [\frac{1}{4}, \frac{3}{4}). \\ x & \text{if } x \in [\frac{3}{4}, 1]. \end{cases}$$

We know that $M(\underline{F}_A, \overline{F}_A) \supseteq M(P_{\Gamma_A}^*)$. Let us now see that for every probability P such that $\underline{F}_A \leq F_P \leq \overline{F}_A$, $P \in M(P_{\Gamma_A}^*)$. Let P be one such probability, and let F_P denote its associated cumulative distribution function. Consider now the measurable map $U(\omega) := F_P^{-1}(\omega)$, where F_P^{-1} denotes the pseudo-inverse of the cumulative distribution function F_P . It trivially holds that $U \in S(\Gamma_A)$, and consequently $P_U \in P(\Gamma_A) \subseteq M(P_{\Gamma_A}^*)$. On the other hand, since $F_P^{-1}(\omega) \leq x$ if and only if $\omega \in [0, F_P(x)]$, F_U and F_P coincide:

$$\begin{aligned} F_U(x) &= P(\{\omega \in [0, 1] \mid U(\omega) \leq x\}) = P(\{\omega \in [0, 1] \mid F_P^{-1}(\omega) \leq x\}) \\ &= P(\{\omega \in [0, 1] \mid \omega \leq F_P(x)\}) = P([0, F_P(x)]) = F_P(x). \end{aligned}$$

Thus, $P = P_U$, and consequently $P \in P(\Gamma_A) \subseteq M(P_{\Gamma_A}^*)$. ♦

The following result gives a sufficient condition for the equality between $M(\underline{F}_A, \overline{F}_A)$ and $M(P_{\Gamma_A}^*)$:

Proposition 5.63 *If the initial space is $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random interval is an IVF-set as in Equation (5.5), where $\mu_A(x) = 0$ for every x , then $M(\underline{F}_A, \overline{F}_A) = M(P_{\Gamma_A}^*)$.*

Proof: Assume there is a probability $P \in M(\underline{F}_A, \overline{F}_A)$ such that for some measurable B it satisfies $P(B) \notin [P_{*\Gamma_A}(B), P_{\Gamma_A}^*(B)]$. We consider two cases: $0 \in B$ and $0 \notin B$.

$0 \notin B$: When $0 \notin B$, it holds that $P_{*\Gamma_A}(B) = 0$:

$$P_{*\Gamma_A}(B) = P(\{\omega \mid \Gamma(\omega) \subseteq B\}) = P(\{\omega \mid [0, \mu_A(\omega)] \subseteq B\}) = 0,$$

since $0 \in \Gamma_A(\omega) \setminus B$ for any ω . Then, it holds that $P(B) > P_{\Gamma_A}^*(B)$. In addition, $P_{\Gamma_A}^*(B) = 1 - P_{*\Gamma_A}(B^c)$, and consequently $P_{*\Gamma_A}(B^c)$ must be strictly positive (otherwise $P(B) > P_{\Gamma_A}^*(B) = 1$ and a contradiction arises). Thus, there exists an interval $[0, x] \subseteq B^c$. Let $\varepsilon = \sup\{x : [0, x] \subseteq B^c\}$, and consider two cases:

- Assume that $\varepsilon = \max\{x : [0, x] \subseteq B^c\}$. Then, since $(\varepsilon, 1] \supseteq B$, it holds that:

$$P(B) \leq P((\varepsilon, 1]) = 1 - F_P(\varepsilon),$$

and consequently:

$$1 - F_P(\varepsilon) \geq P(B) > P_{\Gamma_A}^*(B) = 1 - P_{*\Gamma_A}(B^c).$$

But:

$$P_{\Gamma_A^*}(B^c) = P(\{\omega \mid \Gamma_A(\omega) \subseteq B^c\}) = P(\{\omega \mid \Gamma_A(\omega) \subseteq [0, \varepsilon]\}) = \underline{F}_A(\varepsilon).$$

Thus:

$$1 - F_P(\varepsilon) > 1 - P_{*\Gamma_A}(B^c) = 1 - \underline{F}_A(\varepsilon) \Rightarrow \underline{F}_A(\varepsilon) > F(\varepsilon),$$

and a contradiction arises since $P \notin M(\underline{F}_A, \overline{F}_A)$.

- Assume that $\varepsilon \neq \max\{x : [0, x] \subseteq B^c\}$. Then:

$$P_{*\Gamma_A}(B^c) = P(\{\omega \mid \Gamma_A(\omega) \subseteq B^c\}) = P(\{\omega \mid \Gamma_A(\omega) \subseteq [0, \varepsilon]\}) = P_{*\Gamma_A}([0, \varepsilon]).$$

Moreover:

$$[0, \varepsilon] \subseteq B^c \Rightarrow [\varepsilon, 1] \supseteq B \Rightarrow P([\varepsilon, 1]) \geq P(B).$$

Thus, it holds that

$$P([\varepsilon, 1]) \geq P(B) > 1 - P_{*\Gamma_A}([0, \varepsilon]) = P_{\Gamma_A}^*([\varepsilon, 1]).$$

However, note that $F_P(t) \geq \underline{F}_A(t) = F_{1-\nu_A}(t)$ for any t , and:

$$P([\varepsilon, 1]) = 1 - F_P(t^-) \leq 1 - \underline{F}_A(t^-) = P_{\Gamma_A}^*([\varepsilon, 1]),$$

a contradiction.

$0 \in B$: Note that, since $0 \in B$, $P_{\Gamma_A}^*(B) = 1$:

$$P_{\Gamma_A}^*(B) = P(\{\omega \mid \Gamma(\omega) \cap B \neq \emptyset\}) \geq P(\{\omega \mid \Gamma(\omega) \cap \{0\} \neq \emptyset\}) = 1.$$

Then $P(B) < P_{*\Gamma_A}(B)$. Since $P_{*\Gamma_A}(B) > 0$, there exists $[0, x] \subseteq B$. Define $\varepsilon = \sup\{x : [0, x] \subseteq B\}$ and consider two cases:

- Assume that $\varepsilon = \max\{x : [0, x] \subseteq B\}$. Then, $P(B) \geq P([0, \varepsilon]) = F_P(\varepsilon)$. However:

$$\begin{aligned} P_{*\Gamma_A}(B) &= P(\{\omega \mid \Gamma_A(\omega) \subseteq B\}) = P(\{\omega \mid \Gamma_A(\omega) \subseteq [0, \varepsilon]\}) \\ &= \underline{F}_A(\varepsilon) \leq F_P(\varepsilon) \leq P(B), \end{aligned}$$

a contradiction, because we had assumed that $P_{*\Gamma_A}(B) > P(B)$.

- Assume that $\varepsilon \neq \max\{x : [0, x] \subseteq B\}$. Then $P(B) \geq P([0, \varepsilon))$. Moreover,

$$\begin{aligned} P_{*\Gamma_A}(B) &= P(\{\omega \mid \Gamma_A(\omega) \subseteq B\}) = P(\{\omega \mid \Gamma_A(\omega) \subseteq [0, \varepsilon)\}) \\ &= \underline{F}_A(\varepsilon^-) = F_{1-\nu_A}(\varepsilon^-) \leq F_P(\varepsilon^-) = P([0, \varepsilon)) \leq P(B). \end{aligned}$$

This contradicts the assumption of $P_{\Gamma_A^*}(B) > P(B)$. ■

Another sufficient condition for the equality between $M(P_{\Gamma_A}^*)$ and $M(\underline{F}_A, \overline{F}_A)$ is the strict comonotonicity between μ_A and $1 - \nu_A$, that, as we have seen in Corollary 5.59, is equivalent to the existence of a total order between the intervals $[\mu_A(\omega), 1 - \nu_A(\omega)]$.

Proposition 5.64 *If the initial space is $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random interval is given by an IF-set as in Equation (5.5), where $\Gamma_A(\omega) \leq \Gamma_A(\omega')$ or $\Gamma_A(\omega) \geq \Gamma_A(\omega')$ for any $\omega, \omega' \in \Omega$, then $M(\underline{F}_A, \overline{F}_A) = M(P_{\Gamma_A}^*)$.*

Proof: In [129, Theorem 4.5] it is proven that when the random interval is defined on $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and its bounds are strictly comonotone, then it is possible to define the random interval $\Gamma : [0, 1] \rightarrow \mathcal{P}([0, 1])$ by:

$$\Gamma(\omega) := [U(\omega), V(\omega)],$$

where U and V denote the quantile functions of the lower and upper bounds of Γ_A , respectively, that are defined by:

$$U(\omega) = \inf\{x \in \mathbb{R} : \omega \leq \underline{F}(x)\} \text{ and } V(\omega) = \inf\{x \in \mathbb{R} : \omega \leq \overline{F}(x)\}.$$

This random interval satisfies $P_\Gamma^* = P_{\Gamma_A}^*$, and consequently $M(P_\Gamma^*) = M(P_{\Gamma_A}^*)$ and $M(\underline{F}, \overline{F}) = M(\underline{F}_A, \overline{F}_A)$. Then, in order to prove the equality $M(P_{\Gamma_A}^*) = M(\underline{F}_A, \overline{F}_A)$ it is sufficient to establish the equality between $M(P_\Gamma^*) = M(\underline{F}, \overline{F})$.

Consider now a probability $P \in M(\underline{F}, \overline{F})$, and define W as the quantile function of F_P . Since $\underline{F} \leq F_P \leq \overline{F}$, $W(\omega)$ is bounded by $U(\omega)$ and $V(\omega)$ for any $\omega \in [0, 1]$. Then, W

is a measurable selection of Γ , and its induce probability P_W belongs to $P(\Gamma)$. Moreover, since $P(\Gamma) = M(P_\Gamma^*)$, P_W also belongs to $M(P_\Gamma^*)$.

Thus, $M(P_\Gamma^*) = M(\underline{E}, \overline{F})$, and therefore $M(P_{\Gamma_A}^*) = M(\underline{E}_A, \overline{F}_A)$. ■

One particular situation where the previous result holds is when μ_A is strictly increasing, ν_A is strictly decreasing and $\mu_A(\omega) = \mu_A(\omega')$ if and only if $\nu_A(\omega) = \nu_A(\omega')$.

Finally, we are going to see that the equality between both credal sets also holds when the bounds of the interval are increasing.

Proposition 5.65 *If the initial space is $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random interval is given by an IF-set as in Equation (5.5), where μ_A is increasing and ν_A is decreasing, then $M(\underline{E}_A, \overline{F}_A) = M(P_{\Gamma_A}^*)$.*

Proof: Let P be a probability in $M(\underline{E}_A, \overline{F}_A)$, and we are going to see that there is a measurable selection V such that $P_V = P$, and therefore $M(\underline{E}_A, \overline{F}_A) \subseteq P(\Gamma_A) \subseteq M(P_{\Gamma_A}^*)$. Since μ_A is increasing, there is a countable number of elements $\omega \in (0, 1)$ such that $\mu_A(\omega) > \sup_{\omega' < \omega} \mu_A(\omega')$. Denote this set by N , and consider the function $V : [0, 1] \rightarrow \mathbb{R}$ defined by:

$$V(\omega) = \begin{cases} \inf\{y : \omega \leq P((-\infty, y])\} & \text{if } \omega \in (0, 1) \setminus N. \\ \mu_A(\omega) & \text{otherwise.} \end{cases}$$

Following the same steps than in [129, Proposition 4.1], this function V can be proved to be a measurable selection of Γ_A such that $P_V = P$. Then, we conclude that $M(\underline{E}_A, \overline{F}_A) \subseteq P(\Gamma_A) \subseteq M(P_{\Gamma_A}^*)$, and then we conclude that both credal sets coincide. ■

These results allow us state a number of sufficient conditions for the equality between the three sets of probabilities $P(\Gamma_A)$, $M(P_{\Gamma_A}^*)$ and $M(\underline{E}_A, \overline{F}_A)$.

Corollary 5.66 *Consider the initial space is $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random interval Γ_A given by an IVF-set as in Equation (5.5). Then, the equalities $P(\Gamma_A) = M(P_{\Gamma_A}^*) = M(\underline{E}_A, \overline{F}_A)$ hold if one of the following conditions is satisfied:*

- μ_A is increasing and ν_A is decreasing.
- $\mu_A(\omega) = 0$ for any $\omega \in [0, 1]$.
- μ_A and $1 - \nu_A$ are strictly comonotone, or equivalently, if $\Gamma_A(\omega) \leq \Gamma_A(\omega')$ or $\Gamma_A(\omega') \leq \Gamma_A(\omega)$ for any $\omega, \omega' \in [0, 1]$.

We have seen sufficient conditions under which the p-box defined from the random interval Γ_A contains the same information than the set of measurable selections. Conversely,

there are situations in which, given a p-box, it is possible to define a random interval Γ_A whose associated p-box coincides with the previous one, and that the probabilistic information given by the p-box is the same that the information given by the set of measurable selections.

Proposition 5.67 *Consider a p-box $(\underline{F}, \overline{F})$ defined on $[0, 1]$ such that both \underline{F} and \overline{F} are right-continuous. Then it is possible to define a random interval $\Gamma : [0, 1] \rightarrow \mathcal{P}([0, 1])$ whose associated p-box is $(\underline{F}, \overline{F})$. In addition, if either \underline{F} and \overline{F} are strictly comonotone or $\overline{F}(x) = 1$, then the random interval Γ satisfies $P(\Gamma) = M(\underline{F}, \overline{F})$.*

Proof: Proposition 2.45 assures that $P(\Gamma) = M(P_{\Gamma_A}^*)$. Given the p-box $(\underline{F}, \overline{F})$, define the random interval $\Gamma_A(\omega) = [U(\omega), V(\omega)]$, where U and V are the quantile functions of \underline{F} and \overline{F} , respectively. Then, the p-box associated with Γ_A is given by:

$$\begin{aligned} \underline{F}_A(t) &= F_V(t) = P(\{\omega \in [0, 1] \mid V(\omega) \leq t\}) = \underline{F}(t). \\ \overline{F}_A(t) &= F_U(t) = P(\{\omega \in [0, 1] \mid U(\omega) \leq t\}) = \overline{F}(t). \end{aligned}$$

Since \underline{F} and \overline{F} are right-continuous, U and V are random variables because their cumulative distribution functions are right-continuous. Assume now that \underline{F} and \overline{F} are strictly comonotone. Then, U and V are also strictly comonotone, and following Proposition 5.64, the credal set $P(\Gamma_A)$ coincides with the credal set $M(\underline{F}, \overline{F})$.

Assume that $\overline{F}(x) = 1$. Then, $U = 0$ almost surely. Applying Proposition 5.63, $P(\Gamma_A) = M(\underline{F}, \overline{F})$. \blacksquare

In Corollary 5.60 we have seen that the upper probability induced by the random set Γ_A defined from an IF-set I_A is a possibility measure if and only if μ_A and ν_A are strictly comonotone on the complementary of a null set. In [199], the following result is proved:

Proposition 5.68 ([199, Corollary 17]) *Assume that Ω/\simeq is order complete and let $(\underline{F}, \overline{F})$ be a p-box. Let $\underline{P}_{(\underline{F}, \overline{F})}$ denote the lower probability associated with $(\underline{F}, \overline{F})$ by means of Equation (2.17). Then the natural extension of $\underline{P}_{(\underline{F}, \overline{F})}$ is a possibility measure if and only if either*

(L1) \underline{F} is 0-1 valued,

(L2) $\overline{F}(x) = \overline{F}(x^-)$ for all $x \in \Omega$ that have no immediate predecessor, and

(L3) $\{x \in \Omega \cup \{0-\} : \underline{F}(x) = 1\}$ has a minimum, where $0-$ is a minimum element on Ω ,

or

(U1) \overline{F} is 0-1 valued,

(U2) $\underline{F}(x) = \underline{F}(x^+)$ for all $x \in \Omega$ that have no immediate successor, and

(U3) $\{x \in \Omega \cup \{0-\} : \overline{F}(x) = 0\}$ has a maximum.

In our context, when the initial space is $[0, 1]$, no element in such interval has immediate predecessor or successor. Assume now that the p-box $(\underline{F}_A, \overline{F}_A)$ defined from the random interval Γ_A as in Equations (5.7) and (5.8) is a possibility measure. Note that since \underline{F}_A and \overline{F}_A are right-continuous, (U2) becomes trivial. On the one hand, assume that \underline{F}_A is 0-1 valued. Then, there exists t^* such that $\underline{F}(t) = 1$ for any $t \geq t^*$ and $\underline{F}(t) = 0$ for any $t < t^*$, and by (L3) it is left-continuous. Equivalently:

$$\begin{aligned}\underline{F}(t) &= P(\{\omega \in [0, 1] \mid \Gamma_A(\omega) \subseteq [0, t]\}) = 1 \text{ for any } t \geq t^*. \\ \underline{F}(t) &= P(\{\omega \in [0, 1] \mid \Gamma_A(\omega) \subseteq [0, t]\}) = 0 \text{ for any } t < t^*.\end{aligned}$$

Then, $1 - \nu_A(\omega) = t^*$ for every $\omega \in [0, 1] \setminus N$ for some null set N on $\beta_{[0,1]}$. On the other hand, assume that \overline{F}_A is 0-1 valued. Then, there exists t^* such that $\overline{F}(t) = 1$ for any $t > t^*$ and $\overline{F}(t) = 0$ for any $t \leq t^*$, and by (U2) it is right-continuous. Equivalently:

$$\begin{aligned}\overline{F}(t) &= P(\{\omega \in [0, 1] \mid \Gamma_A(\omega) \cap [0, t] \neq \emptyset\}) = 1 \text{ for any } t \geq t^*. \\ \overline{F}(t) &= P(\{\omega \in [0, 1] \mid \Gamma_A(\omega) \cap [0, t] \neq \emptyset\}) = 0 \text{ for any } t < t^*.\end{aligned}$$

Thus, $\mu_A(\omega) = t^*$ for every $\omega \in [0, 1] \setminus N$ for some null set N on $\beta_{[0,1]}$. We deduce that:

Proposition 5.69 *Consider the initial space $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random interval Γ_A defined from the IVF-set I_A . Consider the p-box $(\underline{F}_A, \overline{F}_A)$ defined in Equations (5.7) and (5.8). If $(\underline{F}_A, \overline{F}_A)$ defines a possibility measure, then there is a null set N on $\beta_{[0,1]}$ and t^* such that either $1 - \nu_A(\omega) = t^*$ for any $\omega \in [0, 1] \setminus N$ or $\mu_A(\omega) = t^*$ for any $\omega \in [0, 1] \setminus N$. In such a case, $P(\Gamma_A) = M(P_{\Gamma_A}^*) = M(\underline{F}_A, \overline{F}_A)$.*

A non-measurable approach

The previous developments assume that the intuitionistic fuzzy set is defined on a probability space and that the functions μ_A, ν_A are measurable with respect to the σ -field we have on this space and the Borel σ -field on $[0, 1]$. Although this is a standard assumption when considering the probabilities associated with fuzzy events, it is arguably done for mathematical convenience only. In this section, we present an alternative approach where we get rid of the measurability assumptions by means of finitely additive probabilities. This allows us to make a clearer link with p -boxes, by means of Walley's notion of natural extension introduced in Definition 2.32.

Consider thus a intuitionistic fuzzy set A defined on a space Ω . If this set is determined by the functions μ_A, ν_A , we can represent it by means of the multi-valued mapping

$\Gamma_A : \Omega \rightarrow [0, 1]$ given by $\Gamma_A(\omega) = [\mu_A(\omega), 1 - \nu_A(\omega)]$. Note that we are not assuming anymore that this multi-valued mapping is strongly measurable, and now our information about the “true” membership function would be given by the set of functions

$$\{\phi : \Omega \rightarrow [0, 1] : \mu_A(\omega) \leq \phi(\omega) \leq 1 - \nu_A(\omega)\}.$$

Now, if we do not assume the measurability of μ_A, ν_A and consider then the field $\mathcal{P}(\Omega)$ of all events in the initial space, we may not be able to model our uncertainty by means of a σ -additive probability measure. However, we can do so by means of a finitely additive probability measure P or more generally by means of an imprecise probability model [205]. Moreover, the notions of lower and upper probabilities can be generalized to that case [132]. If for instance we consider a finitely additive probability P on $\mathcal{P}(\Omega)$, then by an analogous reasoning to that in Section 5.2.1 we obtain that

$$P_\phi(C) \in [P_{\Gamma_{*A}}(C), P_{\Gamma_A}^*(C)] \quad \forall C \subseteq [0, 1],$$

where $P_{\Gamma_A}^*$ is the completely alternating upper probability given by

$$P_{\Gamma_A}^*(C) = P(\{\omega : \Gamma_A(\omega) \cap C \neq \emptyset\})$$

and its conjugate $P_{*\Gamma_A}$ is the completely monotone lower probability given by

$$P_{*\Gamma_A}(C) = P(\{\omega : \emptyset \neq \Gamma_A(\omega) \subseteq C\})$$

for every $C \subseteq [0, 1]$. Then the information about P_ϕ is given by the set of finitely additive probabilities dominated by $P_{\Gamma_A}^*$, and we do not need to make the distinction between $P(\Gamma_A)$ and $M(P_{\Gamma_A}^*)$ as in Section 5.2.1.

The associated p-box is given now by the set of finitely additive distribution functions (that is, monotone and normalized) that lie between \underline{F}_A and \overline{F}_A , where again $\underline{F}_A, \overline{F}_A$ are given by Equations (5.7) and (5.8), respectively.

This set is equivalent to the set of associated finitely additive probability measures that can be determined by natural extension. This can be determined in the following way ([198]): if we denote by \mathcal{H} the field of subsets of $[0, 1]$ generated by the sets $\{[0, x], (x, 1) : x \in [0, 1]\}$, then any set $B \in \mathcal{H}$ is of the form

$$B := [0, x_1] \cup (x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$$

or

$$B := (x_1, x_2] \cup \dots \cup (x_{2n}, x_{2n+1}]$$

for some $n \in \mathbb{N}, x_1 < x_2 < \dots < x_n \in [0, 1]$. It holds that

$$\underline{E}_{\underline{F}, \overline{F}}([0, x_1] \cup (x_2, x_3] \cup \dots \cup (x_n, 1]) = \underline{F}_A(x_1) + \sum_{i=1}^n \max\{0, \underline{F}_A(x_{2i+1}) - \overline{F}_A(x_{2i})\}$$

and

$$\underline{E}_{\underline{F}, \overline{F}}((x_1, x_2] \cup \dots \cup (x_{2n}, x_{2n+1}]) = \sum_{i=0}^n \max\{0, \underline{F}_A(x_{2i+1}) - \overline{F}_A(x_{2i})\} \quad (5.9)$$

and if we consider any $C \subseteq [0, 1]$, then

$$\underline{E}_{\underline{F}, \overline{F}}(C) = \sup_{B \subseteq C, B \in \mathcal{H}} \underline{E}_{\underline{F}, \overline{F}}(B).$$

The upper probability $P_{\Gamma_A}^*$ is determined by $P_{*\Gamma_A}$ using conjugacy.

It can be easily seen that $P_{*\Gamma_A}$ and the natural extension of the p-box $\underline{E}_{\underline{F}, \overline{F}}$ do not coincide in general, even in sets of the form $(x_1, x_2]$:

Example 5.70 Consider the random interval of Example 5.61. We already know that $P_{*\Gamma_A} \left(\left[\frac{1}{4}, \frac{3}{4} \right] \right) = \frac{1}{2}$. Similarly, it can be proved that $P_{*\Gamma_A} \left(\left(\frac{1}{4}, \frac{3}{4} \right] \right) = \frac{1}{2}$. Now, let us use Equation (5.9) to compute $\underline{E}_{\underline{F}, \overline{F}} \left(\left(\frac{1}{4}, \frac{3}{4} \right] \right)$:

$$\underline{E}_{\underline{F}, \overline{F}} \left(\left(\frac{1}{4}, \frac{3}{4} \right] \right) = \max \left\{ 0, \underline{F}_A \left(\frac{3}{4} \right) - \overline{F}_A \left(\frac{1}{4} \right) \right\} = \max \left\{ 0, \frac{1}{2} - \frac{1}{2} \right\} = 0.$$

We conclude that, in general, $P_{*\Gamma_A}$ and $\underline{E}_{\underline{F}, \overline{F}}$ do not coincide even in sets of the form $(x_1, x_2]$. ♦

Our next example shows that $P_{*\Gamma_A}$ and $\underline{E}_{\underline{F}, \overline{F}}$ do not coincide neither when the bounds of the random interval are increasing.

Example 5.71 Consider the random interval defined by:

$$\Gamma_A(\omega) = \begin{cases} [\omega, 2\omega] & \text{if } \omega \in [0, \frac{1}{3}]. \\ [\frac{1}{3}, \frac{2}{3}] & \text{if } \omega \in (\frac{1}{3}, \frac{2}{3}]. \\ [2\omega - 1, \omega] & \text{otherwise.} \end{cases}$$

The bounds of its associated p-box are defined by:

$$\underline{F}_A(x) = \begin{cases} \frac{1}{2}x & \text{if } x \in [0, \frac{2}{3}]. \\ x & \text{otherwise.} \end{cases}$$

$$\overline{F}_A(x) = \begin{cases} x & \text{if } x \in [0, \frac{1}{3}]. \\ \frac{1}{2}x + \frac{1}{2} & \text{otherwise.} \end{cases}$$

Then, $P_{*\Gamma_A} \left(\left[\frac{1}{3}, \frac{2}{3} \right] \right) = \frac{1}{3}$. However, it holds that:

$$\underline{E}_{\underline{F}_A, \overline{F}_A} \left(\left(\frac{1}{3}, \frac{2}{3} \right] \right) = \underline{F}_A \left(\frac{2}{3} \right) - \overline{F}_A \left(\frac{1}{3} \right) = \frac{2}{3} - \frac{2}{3} = 0.$$

Furthermore:

$$\underline{E}_{\underline{F}_A, \overline{F}_A} \left(\left[\frac{1}{3}, \frac{2}{3} \right] \right) = \sup_{B \subseteq [\frac{1}{3}, \frac{2}{3}], B \in \mathcal{H}} \underline{E}_{\underline{F}_A, \overline{F}_A}(B) \leq \underline{E}_{\underline{F}_A, \overline{F}_A} \left(\left(\frac{1}{3}, \frac{2}{3} \right) \right) = 0.$$

Thus, the natural extension is less informative than the original lower probability. \blacklozenge

Next we show that the lower probability and the natural extension defined of the p-box coincide when $\mu_A = 0$.

Proposition 5.72 Consider the initial space $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ and the random interval defined from an IF-set A with $\mu_A = 0$. Then, $\underline{E}_{\underline{F}_A, \overline{F}_A} = P_{*\Gamma_A}$.

Proof: We know that $\mu_A = 0$ implies that $\overline{F}_A = 1$. Let us prove the equality between the natural extension and the lower probability following several steps:

1. Let B be a set on \mathcal{H} . We have several cases:

- Assume that $B = [0, x]$. Then:

$$\begin{aligned} P_{*\Gamma_A}([0, x]) &= P(\{\omega : \Gamma(\omega) \subseteq [0, x]\}) = \underline{F}_A(x). \\ \underline{E}_{\underline{F}_A, \overline{F}_A}([0, x]) &= \underline{F}_A(x). \end{aligned}$$

- Assume now that $B = [0, x_1] \cup [x_2, x_3] \cup \dots \cup [x_{2k}, x_{2k+1}]$, with $x_1 < x_2 < \dots < x_n$. Then:

$$\begin{aligned} P_{*\Gamma_A}(B) &= P(\{\omega : \Gamma_A(\omega) \subseteq B\}) = P(\{\omega : \Gamma_A(\omega) \subseteq [0, x_1]\}) = \underline{F}_A(x_1). \\ \underline{E}_{\underline{F}_A, \overline{F}_A}(B) &= \underline{F}_A(x_1) + \sum_{i=1}^n \max\{0, \underline{F}_A(x_{2i+1}) - \overline{F}_A(x_{2i})\} \\ &= \underline{F}_A(x_1) + \sum_{i=1}^n \max\{0, \underline{F}_A(x_{2i+1}) - 1\} = \underline{F}_A(x_1). \end{aligned}$$

- Finally, assume that $B = (x_1, x_2] \cup \dots \cup (x_{2n}, x_{2n+1}]$, with $x_1 < x_2 < \dots < x_n$. Then:

$$\begin{aligned} P_{*\Gamma_A}(B) &= P(\{\omega : \Gamma_A(\omega) = [0, 1 - \nu_A(\omega)] \subseteq B\}) = 0. \\ \underline{E}_{\underline{F}_A, \overline{F}_A}(B) &= \sum_{i=1}^n \max\{0, \underline{F}_A(x_{2i+1}) - \overline{F}_A(x_{2i})\} \\ &= \sum_{i=1}^n \max\{0, \underline{F}_A(x_{2i+1}) - 1\} = 0. \end{aligned}$$

Then, $\underline{E}_{\underline{F}_A, \overline{F}_A}$ and $P_{*\Gamma_A}$ coincide for elements in \mathcal{H} .

2. Consider $C \subseteq [0, 1]$. Denote by $x^* = \sup\{x : [0, x] \subseteq C\}$. We have several cases:

- Assume that $\{x : [0, x] \subseteq C\} = \emptyset$, that means that $0 \notin C$. Then, $0 \notin B$ for every $B \in \mathcal{H}$, and then $\underline{E}_{\underline{F}_A, \overline{F}_A}(B) = 0$. Thus, we conclude that

$$\underline{E}_{\underline{F}_A, \overline{F}_A}(C) = \sup_{B \subseteq C, B \in \mathcal{H}} (B) = 0.$$

Furthermore, since $0 \notin C$, $P_{*\Gamma_A}(C) = 0$.

- Now, assume that $x^* = \max\{x : [0, x] \subseteq C\}$, that means that $0 \in C$ and there is x^* such that $[0, x^*] \subseteq C$ but $[0, x^* + \varepsilon] \not\subseteq C$ for any $\varepsilon > 0$. Then:

$$\begin{aligned} P_{*\Gamma_A}(C) &= P(\{\omega : \Gamma_A(\omega) \subseteq C\}) = P(\{\omega : \Gamma_A(\omega) \subseteq [0, x^*]\}) = \underline{E}_A(x^*). \\ \underline{E}_{\underline{F}_A, \overline{F}_A}([0, x^*]) &= \underline{E}_A(x^*). \end{aligned}$$

Furthermore, as in the previous case:

$$\underline{E}_{\underline{F}_A, \overline{F}_A}([0, x^*]) = \underline{E}_{\underline{F}_A, \overline{F}_A}(B)$$

for any B such that $[0, x^*] \subseteq B$, and consequently

$$\underline{E}_{\underline{F}_A, \overline{F}_A}(C) = \underline{E}_{\underline{F}_A, \overline{F}_A}([0, x^*]) = \underline{E}_A(x^*).$$

- Finally, assume that x^* is a supremum, not a maximum, that is: $[0, x^*) \subseteq C$ but $x^* \notin C$. Then:

$$\begin{aligned} P_{*\Gamma_A}(C) &= P(\{\omega : \Gamma_A(\omega) \subseteq [0, x^*)\}) = \lim_{\varepsilon \rightarrow 0} P(\{\omega : 1 - \nu_A(\omega) \leq x^* - \varepsilon\}) \\ &= \lim_{\varepsilon \rightarrow 0} \underline{E}_A(x^* - \varepsilon) = \lim_{\varepsilon \rightarrow 0} P_{*\Gamma_A}([0, x^* - \varepsilon]) \\ &= \lim_{\varepsilon \rightarrow 0} \underline{E}_{\underline{F}_A, \overline{F}_A}([0, x^* - \varepsilon]) \\ &= \sup_{B \subseteq [0, x^*), B \in \mathcal{H}} \underline{E}_{\underline{F}_A, \overline{F}_A}(B) = \underline{E}_{\underline{F}_A, \overline{F}_A}([0, x^*)). \end{aligned}$$

In addition, every $B \in \mathcal{H}$ such that $[0, x^*) \subseteq B$ satisfies that $\underline{E}_{\underline{F}_A, \overline{F}_A}(B) = \underline{E}_{\underline{F}_A, \overline{F}_A}([0, x^*)$. Then, the lower probability and the natural extension coincide. ■

We could think that the lower probability and the natural extension of the associated p-box also coincide when the bounds of the random interval are strictly comonotone functions. However, we can find examples where such equality does not hold.

Example 5.73 Consider the random interval Γ_A defined on $([0, 1], \beta_{[0,1]}, \lambda_{[0,1]})$ by:

$$\Gamma_A(\omega) = \begin{cases} [\frac{1}{2} - \omega, 1 - \omega] & \text{if } \omega \in [0, \frac{1}{4}]. \\ [\frac{1}{4}, \frac{3}{4}] & \text{if } \omega \in [\frac{1}{4}, \frac{3}{4}]. \\ [\omega - \frac{1}{2}, \omega] & \text{if } \omega \in [\frac{3}{4}, 1]. \end{cases}$$

Since $\mu_A(\omega) = (1 - \nu_A(\omega)) - \frac{1}{2}$, we see that μ_A and $1 - \nu_A$ are strictly comonotone. Its associated p-box is defined by:

$$\bar{F}_A(t) = \begin{cases} 0 & \text{if } t \in [0, \frac{1}{4}), \\ 2t & \text{if } t \in [\frac{1}{4}, \frac{1}{2}], \\ 1 & \text{if } t \in (\frac{1}{2}, 1], \end{cases} \text{ and } \underline{F}_A(t) = \begin{cases} 0 & \text{if } t \in [0, \frac{3}{4}], \\ 2t - 1 & \text{if } t \in (\frac{3}{4}, 1]. \end{cases}$$

Let us compute $P_{*\Gamma_A}$ and $\underline{E}_{\underline{F}_A, \bar{F}_A}$ for the set $(\frac{1}{4}, \frac{7}{8}]$:

$$P_{*\Gamma_A} \left(\left(\frac{1}{4}, \frac{7}{8} \right] \right) = P \left(\left\{ \omega : \Gamma_A(\omega) \subseteq \left(\frac{1}{4}, \frac{7}{8} \right] \right\} \right) = \frac{1}{4}.$$

$$\underline{E}_{\underline{F}_A, \bar{F}_A} \left(\left(\frac{1}{4}, \frac{7}{8} \right] \right) = \max \left\{ 0, \underline{F}_A \left(\frac{1}{4} \right) - \bar{F}_A \left(\frac{7}{8} \right) \right\} = \frac{1}{4}.$$

Thus, they coincide. However, we are going to check that they do not agree on the set $[\frac{1}{4}, \frac{7}{8}]$.

$$P_{*\Gamma_A} \left(\left[\frac{1}{4}, \frac{7}{8} \right] \right) = P \left(\left\{ \omega : \Gamma_A(\omega) \subseteq \left[\frac{1}{4}, \frac{7}{8} \right] \right\} \right) = \frac{3}{4}.$$

By definition, $\underline{E}_{\underline{F}_A, \bar{F}_A} \left(\left[\frac{1}{4}, \frac{7}{8} \right] \right) = \sup_{B \subseteq [\frac{1}{4}, \frac{7}{8}], B \in \mathcal{H}} \underline{E}_{\underline{F}_A, \bar{F}_A}(B)$. But

$$\underline{E}_{\underline{F}_A, \bar{F}_A}(B) \leq \underline{E}_{\underline{F}_A, \bar{F}_A} \left(\left(\frac{1}{4}, \frac{7}{8} \right] \right) = \frac{1}{4}$$

for any $B \subseteq [\frac{1}{4}, \frac{7}{8}]$ in \mathcal{H} . Thus,

$$P_{*\Gamma_A} \left(\left[\frac{1}{4}, \frac{7}{8} \right] \right) > \underline{E}_{\underline{F}_A, \bar{F}_A} \left(\left[\frac{1}{4}, \frac{7}{8} \right] \right). \blacklozenge$$

5.2.2 Connection with other approaches

We now investigate the connection between the framework we have presented and other theories that can be found in the literature. For this aim, we first investigate the connection with the approach of Grzegorzewski and Mrowka ([86]) and then we establish a one-to-one relationship between IVF-sets, p-boxes and clouds.

Probabilities associated with IF-Sets

One of the most important works on the connection between IF-sets and imprecise probabilities is the work carried out in [86] on the probabilities of IF-sets. Given a probability space (Ω, \mathcal{A}, P) , the probability associated with an IF-set A is a number of the interval

$$\left[\int_{\Omega} \mu_A dP, \int_{\Omega} 1 - \nu_A dP \right]. \quad (5.10)$$

Using this definition, in [86] a link is established with probability theory by considering the appropriate operators in the spaces of real intervals and of intuitionistic fuzzy sets. Note that in this work it is assumed that we have a structure of probability space on Ω and that moreover the functions μ_A, ν_A are measurable, as we have done in Section 5.2.1.

Remark 5.74 *This definition generalises an earlier definition by Zadeh [215] for fuzzy events. He defined the probability of a fuzzy event μ_A by:*

$$P(\mu_A) = \int_{\Omega} \mu_A dP = E[\mu_A].$$

Although Zadeh proved that this definition satisfies the axioms of Kolmogorov when considering the minimum operator for making intersections, it was proved in [144] that this does not happen for any t -norm (see [100] for a complete review on t -norms). In fact, it was proved that every strict and continuous t -norm made Zadeh's probability to satisfy Kolmogorov axioms, while the Łukasiewicz operator is the only nilpotent and continuous t -norm that satisfies these axioms. ♦

If we consider the random interval associated with the intuitionistic fuzzy set A in Equation (5.5), we can see that the interval in Equation (5.10) corresponds simply to the set of expectations of the measurable selections of Γ_A : it follows from [130, Theorem 14] that if we consider the mapping $id : [0, 1] \rightarrow [0, 1]$, then the *Aumann integral* [13] of $(id \circ \Gamma_A)$, defined on Equation (2.26), satisfies that

$$\left[\inf(A) \int (id \circ \Gamma_A) dP, \sup(A) \int (id \circ \Gamma_A) dP \right] = \left[(C) \int id dP_{\Gamma_A}^*, (C) \int id dP_{*\Gamma_A} \right],$$

where (C) is used to denote the Choquet integral [39, 60] with respect to the non-additive measures $P_{*\Gamma_A}, P_{\Gamma_A}^*$, respectively. Since on the other hand it is immediate to see that

$$\sup(A) \int (id \circ \Gamma_A) dP = \int (1 - \nu_A) dP$$

and

$$\inf(A) \int (id \circ \Gamma_A) dP = \int \mu_A dP,$$

we deduce that the probabilistic information about the intuitionistic fuzzy set A can be determined in particular by the lower and upper probabilities of its associated random interval. Note moreover that the Aumann integral of a random set is not convex in general, and it is only guaranteed to be so when the probability space (Ω, \mathcal{A}, P) is non-atomic.

A one-to-one relationship between p-boxes and IFS

In Section 5.2.1, we saw that the correspondence between interval-valued fuzzy sets and p-boxes on $[0, 1]$ is many-to-one, in the sense that many different IFS determine the same

lower and upper distribution functions. In this section, we consider a subset of the class of IFS for which a bijection can be established with the set of p-boxes. In contradistinction to our work in Section 5.2.1, the p-box we shall establish here shall be established in the possibility space Ω , that we shall consider here to be the unit interval.

Denote by $IF^*([0, 1])$ the set:

$$IF^*([0, 1]) = \{A \in IFSS(\Omega) \mid \mu_A \text{ increasing and } \nu_A \text{ decreasing}\}.$$

Denote also $\mathcal{F}([0, 1])$ the set of all p-boxes on $[0, 1]$, and let us define the correspondences:

$$\begin{aligned} f_1 : \mathcal{F}([0, 1]) &\longrightarrow IF^*([0, 1]) \\ (\underline{F}, \overline{F}) &\longmapsto A_{(\underline{F}, \overline{F})} = \langle (x, \underline{F}(x), 1 - \overline{F}(x)) \rangle \\ \\ f_2 : IF^*([0, 1]) &\longrightarrow \mathcal{F}([0, 1]) \\ A &\longmapsto (\mu_A, 1 - \nu_A) \end{aligned}$$

We can see that every IFS A has an associated p-box: $(\mu_A, 1 - \nu_A)$. The interpretation here would be that $(\mu_A, 1 - \nu_A)$ models the imprecise information about the distribution function associated with the set A , instead of about the membership function, as we did in Section 5.2.1.

The following properties follow immediately, and therefore their proof is omitted:

Proposition 5.75 *Let f_1, f_2 be the two correspondences between $\mathcal{F}([0, 1])$ and $IF^*([0, 1])$ considered above. Then:*

- (a) f_1, f_2 are bijective, and $f_1 = f_2^{-1}$.
- (b) $f_1((\underline{F}, \overline{F})) \in \Gamma \Leftrightarrow \underline{F} = \overline{F}$.
- (c) $f_2(A) = (F, F) \Leftrightarrow A \in FS(\Omega)$.

Another property assures that there exists a relationship between application f_1 and the stochastic order $(FSD_{2,5})$:

$$(\underline{F}_1, \overline{F}_1) \succeq_{FSD_{2,5}} (\underline{F}_2, \overline{F}_2) \Leftrightarrow f_1((\underline{F}_1, \overline{F}_1)) \subseteq f_1((\underline{F}_2, \overline{F}_2)).$$

A one-to-one relationship between clouds and IFS

A similar correspondence can be made between intuitionistic fuzzy sets and clouds. Recall that a cloud is a pair of functions δ, π such that $\delta \leq \pi$ and there are $x, y \in [0, 1]$ such that $\delta(x) = 0$ and $\pi(y) = 1$. Let us denote by IF_* the following set:

$$IF_* = \{A \in IFSS(\Omega) \mid \mu_A(x) = 0 \text{ and } \nu_A(y) = 0 \text{ for some } x, y \in [0, 1]\}.$$

Then, if we denote by $Cl([0, 1])$ the set of all the clouds on $[0, 1]$, the following functions can be defined:

$$\begin{aligned} g_1 : \quad Cl([0, 1]) &\longrightarrow IF_*([0, 1]) \\ (\delta, \pi) &\mapsto A_{(\delta, \pi)} = \langle (x, \delta(x), 1 - \pi(x)) \rangle \\ \\ g_2 : \quad IF_*([0, 1]) &\longrightarrow Cl([0, 1]) \\ A &\mapsto (\mu_A, 1 - \nu_A) \end{aligned}$$

A cloud (δ, π) is called *thin* ([168]), when $\delta = \pi$; in that case, its associated IVF-sets by g_1 becomes $\langle (x, \delta, 1 - \delta) \rangle \in FS(\Omega)$, that is, a fuzzy set.

This is consistent in the sense that, given a possibility distribution π , it has an associated fuzzy sets $\mu(x) := \pi(x)$. Thus, this is a more general approach that contains the relationship between fuzzy sets and possibility distribution as a particular case.

Another particular type of clouds are the *fuzzy clouds*, for which $\delta = 0$. In such a case the associated IFS is $\langle (x, 0, 1 - \pi) \rangle$.

Some immediate properties of the above correspondences are the following:

Proposition 5.76 *Let g_1, g_2 be the correspondences between $Cl([0, 1])$ and $IF_*([0, 1])$ considered above. The following conditions hold:*

- (a) $g_1((\delta, \pi)) \in FS(\Omega) \Leftrightarrow \delta = \pi \Leftrightarrow (\delta, \pi)$ is a thin cloud.
- (b) $g_2(A) = (\delta, \delta) \Leftrightarrow A \in FS(\Omega)$.
- (c) g_1, g_2 are bijective, and $g_1 = g_2^{-1}$.

The above correspondence is related to the connection between clouds and imprecise probabilities established in [65], where the credal set associated with a cloud (δ, π) is the set of probability measures on Ω satisfying $\mathcal{M}((\pi, 1 - \delta)) = \mathcal{M}(\pi) \cap \mathcal{M}(1 - \delta)$, where $\mathcal{M}(\pi)$ (respectively $\mathcal{M}(1 - \delta)$) is the credal set associated with the possibility measure π (respectively $1 - \delta$).

5.3 Applications

In the previous sections we have presented a theoretical study of comparison measures for intuitionistic fuzzy sets, focusing in the study of IF-divergences, and we have also investigated the connection between IVF-sets and imprecise probabilities.

Now we shall present some possible applications of the theories we have developed. On one hand, we will see how IF-divergences can be applied in multiple attribute decision

making ([211]), and we will outline some examples of application in pattern recognition ([92, 93, 114]). On the other hand, we shall see how the connection between IVF-sets and imprecise probabilities allows us to extend stochastic dominance to the comparison more than two sets of cumulative distribution functions.

5.3.1 Application to pattern recognition

One interesting area of application of comparison measures between IF-sets is in pattern recognition ([92, 93, 114]). Let us consider a universe $\Omega = \{\omega_1, \dots, \omega_n\}$, and assume the patterns A_1, \dots, A_m , that are represented by IF-sets. Then:

$$A_j = \{(\omega_i, \mu_{A_j}(\omega_i), \nu_{A_j}(\omega_i) \mid i = 1, \dots, n\}, \text{ for } j = 1, \dots, m.$$

If B is a sample that is also represented by an IF-set, and we want to classify it into one of the patterns, we can measure the difference between B and A_i :

$$D_{\text{IFS}}(A_1, B), \dots, D_{\text{IFS}}(A_m, B),$$

where D_{IFS} can be an IF-divergence or an IF-dissimilarity. Finally, we associate B to the pattern A_j whenever $D_{\text{IFS}}(A_j, B) = \min_{i=1, \dots, m} (D_{\text{IFS}}(A_i, B))$, i.e., we classify B into the pattern from which it differs the least.

Example 5.77 ([114, Section 4]) Consider a possibility space with three elements, $\Omega = \{\omega_1, \omega_2, \omega_3\}$, and the following three patterns:

$$\begin{aligned} A_1 &= \{(\omega_1, 0.1, 0.1), (\omega_2, 0.5, 0.4), (\omega_3, 0.1, 0.9)\}. \\ A_2 &= \{(\omega_1, 0.5, 0.5), (\omega_2, 0.7, 0.3), (\omega_3, 0, 0.8)\}. \\ A_3 &= \{(\omega_1, 0.7, 0.2), (\omega_2, 0.1, 0.8), (\omega_3, 0.4, 0.4)\}. \end{aligned}$$

Assume that a sample $B = \{(\omega_1, 0.4, 0.4), (\omega_2, 0.6, 0.2), (\omega_3, 0, 0.8)\}$ is given, and let us consider the Hamming and the Hausdorff distances for IF-sets. We obtain the following results.

$$\begin{aligned} l_{\text{IFS}}(A_1, B) &= 1, & l_{\text{IFS}}(A_2, B) &= 0.4, & l_{\text{IFS}}(A_3, B) &= 1.3, \\ d_{\text{H}}(A_1, B) &= 0.6, & d_{\text{H}}(A_2, B) &= 0.2, & d_{\text{H}}(A_3, B) &= 1.3. \end{aligned}$$

Thus, both distances classify B into the pattern A_2 , because

$$\begin{aligned} l_{\text{IFS}}(A_2, B) &\leq l_{\text{IFS}}(A_1, B), l_{\text{IFS}}(A_3, B). \\ d_{\text{H}}(A_2, B) &\leq d_{\text{H}}(A_1, B), d_{\text{H}}(A_3, B). \blacklozenge \end{aligned}$$

In the framework of pattern recognition it is usually assumed that every point u_i in the universe has the same weight, that is, $\alpha_i = \frac{1}{n}$ for $i = 1, \dots, n$. However, it is possible that the weight vector $\alpha = (\alpha_1, \dots, \alpha_n)$ is not constant, that is, $\alpha_i \geq 0$ for $i = 1, \dots, n$ and $\alpha_1 + \dots + \alpha_n = 1$.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$\mu_{C_1}(\omega_i)$	0.739	0.033	0.188	0.492	0.020	0.739
$\nu_{C_1}(\omega_i)$	0.125	0.818	0.626	0.358	0.628	0.125
$\mu_{C_2}(\omega_i)$	0.124	0.030	0.048	0.136	0.019	0.393
$\nu_{C_2}(\omega_i)$	0.665	0.825	0.800	0.648	0.823	0.653
$\mu_{C_3}(\omega_i)$	0.449	0.662	1.000	1.000	1.000	1.000
$\nu_{C_3}(\omega_i)$	0.387	0.298	0.000	0.000	0.000	0.000
$\mu_{C_4}(\omega_i)$	0.280	0.521	0.470	0.295	0.188	0.735
$\nu_{C_4}(\omega_i)$	0.715	0.368	0.423	0.658	0.806	0.118
$\mu_{C_5}(\omega_i)$	0.326	1.000	0.182	0.156	0.049	0.675
$\nu_{C_5}(\omega_i)$	0.452	0.000	0.725	0.765	0.896	0.263
$\mu_B(\omega_i)$	0.629	0.524	0.210	0.218	0.069	0.658
$\nu_B(\omega_i)$	0.303	0.356	0.689	0.753	0.876	0.256

Table 5.2: Six kinds of materials are represented by IF-sets.

To deal with this situation, we propose the following method. Let us consider a local IF-divergence D_{IFS} , and for every point u_i let us compute the following:

$$D_{\text{IFS}}(A_j, B) - D_{\text{IFS}}(A_j \cup \{\omega_i\}, B \cup \{\omega_i\}) = h_{\text{IFS}}(\mu_{A_j}(\omega_i), \nu_{A_j}(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)).$$

Then, for every $j \in \{1, \dots, m\}$ we have that

$$\begin{aligned} d(A_j, B) &= \sum_{i=1}^n \omega_i (D_{\text{IFS}}(A_j, B) - D_{\text{IFS}}(A_j \cup \{\omega_i\}, B \cup \{\omega_i\})) \\ &= \sum_{i=1}^n \alpha_i h_{\text{IFS}}(\mu_{A_j}(\omega_i), \nu_{A_j}(\omega_i), \mu_B(\omega_i), \nu_B(\omega_i)). \end{aligned}$$

Then, we classify the sample B into the pattern A_j if

$$d(A_j, B) = \min_{i=1, \dots, m} (d(A_i, B)).$$

Example 5.78 ([206, Example 4.2]) Consider five kinds of mineral fields, each of them featured by the content of six minerals and containing one kind of typical hybrid mineral. The five kinds of typical hybrid mineral are represented by IF-sets C_1 , C_2 , C_3 , C_4 and C_5 in $\Omega = \{\omega_1, \dots, \omega_6\}$, respectively. Assume that we are given another kind of hybrid mineral B , and that we want to classify it into one of the aforementioned mineral fields. Assume that the IF-sets C_i and B are defined in Table 5.2, and that our experts have established the following weight vector on Ω : $\alpha = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\}$. Let us use our method to classify B . If we consider the Hamming distance for IF-sets as local

IF-divergence, we obtain that:

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$l_{\text{IFS}}(C_1, B) - l_{\text{IFS}}(C_1 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.178	0.491	0.085	0.395	0.297	0.131
$l_{\text{IFS}}(C_2, B) - l_{\text{IFS}}(C_2 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.505	0.494	0.162	0.187	0.103	0.397
$l_{\text{IFS}}(C_3, B) - l_{\text{IFS}}(C_3 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.180	0.138	0.790	0.782	0.931	0.342
$l_{\text{IFS}}(C_4, B) - l_{\text{IFS}}(C_4 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.412	0.012	0.266	0.095	0.119	0.138
$l_{\text{IFS}}(C_5, B) - l_{\text{IFS}}(C_5 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.303	0.476	0.036	0.062	0.020	0.024

whence

$$d(C_1, B) = \frac{1}{4}0.178 + \frac{1}{4}0.491 + \frac{1}{8}0.085 + \frac{1}{8}0.395 + \frac{1}{8}0.297 + \frac{1}{8}0.131 = 0.2808.$$

$$d(C_2, B) = \frac{1}{4}0.505 + \frac{1}{4}0.494 + \frac{1}{8}0.162 + \frac{1}{8}0.187 + \frac{1}{8}0.103 + \frac{1}{8}0.397 = 0.3559.$$

$$d(C_3, B) = \frac{1}{4}0.180 + \frac{1}{4}0.138 + \frac{1}{8}0.790 + \frac{1}{8}0.782 + \frac{1}{8}0.931 + \frac{1}{8}0.342 = 0.4351.$$

$$d(C_4, B) = \frac{1}{4}0.412 + \frac{1}{4}0.012 + \frac{1}{8}0.266 + \frac{1}{8}0.095 + \frac{1}{8}0.119 + \frac{1}{8}0.138 = 0.1833.$$

$$d(C_5, B) = \frac{1}{4}0.303 + \frac{1}{4}0.476 + \frac{1}{8}0.036 + \frac{1}{8}0.062 + \frac{1}{8}0.020 + \frac{1}{8}0.024 = 0.2125.$$

Thus, we classify B into the hybrid mineral C_4 .

If we repeat the process with local IF-divergence d_H , we obtain the following:

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$d_H(C_1, B) - d_H(C_1 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.178	0.491	0.063	0.395	0.248	0.131
$d_H(C_2, B) - d_H(C_2 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.505	0.494	0.162	0.105	0.053	0.397
$d_H(C_3, B) - d_H(C_3 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.180	0.138	0.790	0.782	0.931	0.342
$d_H(C_4, B) - d_H(C_4 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.412	0.012	0.266	0.095	0.119	0.138
$d_H(C_5, B) - d_H(C_5 \cup \{\omega_i\}, B \cup \{\omega_i\})$	0.303	0.476	0.036	0.062	0.020	0.017

Then:

$$d(C_1, B) = \frac{1}{4}0.178 + \frac{1}{4}0.491 + \frac{1}{8}0.063 + \frac{1}{8}0.395 + \frac{1}{8}0.248 + \frac{1}{8}0.131 = 0.2719.$$

$$d(C_2, B) = \frac{1}{4}0.505 + \frac{1}{4}0.494 + \frac{1}{8}0.162 + \frac{1}{8}0.105 + \frac{1}{8}0.053 + \frac{1}{8}0.397 = 0.3394.$$

$$d(C_3, B) = \frac{1}{4}0.180 + \frac{1}{4}0.138 + \frac{1}{8}0.790 + \frac{1}{8}0.782 + \frac{1}{8}0.931 + \frac{1}{8}0.342 = 0.4351.$$

$$d(C_4, B) = \frac{1}{4}0.412 + \frac{1}{4}0.012 + \frac{1}{8}0.266 + \frac{1}{8}0.095 + \frac{1}{8}0.119 + \frac{1}{8}0.138 = 0.1833.$$

$$d(C_5, B) = \frac{1}{4}0.303 + \frac{1}{4}0.476 + \frac{1}{8}0.036 + \frac{1}{8}0.062 + \frac{1}{8}0.020 + \frac{1}{8}0.017 = 0.2116,$$

and we conclude that we also should classify B into the hybrid mineral C_4 .♦

5.3.2 Application to decision making

In [211], Xu showed how measures of similarity for IF-sets (and, consequently, also IF-dissimilarities) can be applied within multiple attribute decision making. Let us overview the main aspects of this application.

We use the following notation: let $A = \{A_1, \dots, A_m\}$ denote a set of m alternatives, let $C = \{C_1, \dots, C_n\}$ be a set of attributes and let $\alpha = \{\alpha_1, \dots, \alpha_n\}$ be its associated weight vector (i.e., it holds that $\alpha_i \geq 0$ for every $i = 1, \dots, n$ and that $\alpha_1 + \dots + \alpha_n = 1$).

Every alternative A_i can be represented by means of an IF-set:

$$A_i = \{(C_j, \mu_{A_i}(C_j), \nu_{A_i}(C_j)) \mid j = 1, \dots, n\}.$$

Thus, $\mu_{A_i}(C_j)$ and $\nu_{A_i}(C_j)$ stand for the degree in which alternative A_i agrees and does not agree with characteristic C_j , respectively.

Xu ([211]) defined the IF-sets A^+ and A^- in the following way:

$$\begin{aligned} A^+ &= \{(C_j, \mu_{A^+}(C_j), \nu_{A^+}(C_j)) \mid j = 1, \dots, n\} \text{ and} \\ A^- &= \{(C_j, \mu_{A^-}(C_j), \nu_{A^-}(C_j)) \mid j = 1, \dots, n\}, \end{aligned}$$

where

$$\mu_{A^+}(C_j) = \max_{i=1, \dots, m} (\mu_{A_i}(C_j)), \quad \nu_{A^+}(C_j) = \min_{i=1, \dots, m} (\nu_{A_i}(C_j)), \quad (5.11)$$

$$\mu_{A^-}(C_j) = \min_{i=1, \dots, m} (\mu_{A_i}(C_j)), \quad \nu_{A^-}(C_j) = \max_{i=1, \dots, m} (\nu_{A_i}(C_j)), \quad (5.12)$$

that is, $A^+ = \bigcup_{i=1}^m A_i$ and $A^- = \bigcap_{i=1}^m A_i$.

These IF-sets can be interpreted as the ‘‘optimal’’ and the ‘‘least optimal’’ alternatives. Therefore, the preferred alternative in A would be the one that is simultaneously more similar to A^+ and more different to A^- .

In order to measure how different is A_i to both A^+ and A^- , Xu considered some different functions, such as:

$$\begin{aligned} D(A^+, A_i) &= \left[\sum_{j=1}^n \alpha_j (|\mu_{A^+}(C_j) - \mu_{A_i}(C_j)|^\beta + |\nu_{A^+}(C_j) - \nu_{A_i}(C_j)|^\beta \right. \\ &\quad \left. + |\pi_{A^+}(C_j) - \pi_{A_i}(C_j)|^\beta \right]^{\frac{1}{\beta}} \end{aligned}$$

and

$$\begin{aligned} D(A^-, A_i) &= \left[\sum_{j=1}^n \alpha_j (|\mu_{A^-}(C_j) - \mu_{A_i}(C_j)|^\beta + |\nu_{A^-}(C_j) - \nu_{A_i}(C_j)|^\beta \right. \\ &\quad \left. + |\pi_{A^-}(C_j) - \pi_{A_i}(C_j)|^\beta \right]^{\frac{1}{\beta}}. \end{aligned}$$

Besides, Xu consider the quotient:

$$d_i = \frac{D(A^+, A_i)}{D(A^+, A_i) + D(A^-, A_i)}.$$

Then, the greater the value d_i , the better the alternative A_i .

Next we propose a modification of the above method. Let us consider a local IF-divergence D_{IFS} , so that for every pair of IF-sets A and B , $D_{\text{IFS}}(A, B)$ can be expressed by:

$$D_{\text{IFS}}(A, B) = \sum_{i=1}^n h_{\text{IFS}}(\mu_A(C_i), \nu_A(C_i), \mu_B(C_i), \nu_B(C_i)).$$

We consider the IF-set A_i , that represents the i -th alternative, and for every $j \in \{1, \dots, n\}$ we compute the following:

$$D_{\text{IFS}}(A^+, A_i) - D_{\text{IFS}}(A^+ \cup \{C_j\}, A_i \cup \{C_j\}) = h_{\text{IFS}}(\mu_{A^+}(C_j), \nu_{A^+}(C_j), \mu_{A_i}(C_j), \nu_{A_i}(C_j)).$$

This quantity measures how different A^+ and A_i are with respect to element C_j . Then, we can compute the difference between A_i and A^+ :

$$d(A_i, A^+) = \sum_{j=1}^n \alpha_j h_{\text{IFS}}(\mu_{A^+}(C_j), \nu_{A^+}(C_j), \mu_{A_i}(C_j), \nu_{A_i}(C_j)).$$

In this way $d(A_i, A^+)$ measures how much difference there is between A_i and the optimal set A^+ .

Similarly, we can compute the difference between A_i and A^- :

$$d(A_i, A^-) = \sum_{j=1}^n \alpha_j h_{\text{IFS}}(\mu_{A^-}(C_j), \nu_{A^-}(C_j), \mu_{A_i}(C_j), \nu_{A_i}(C_j)).$$

Thus, $d(A_i, A^-)$ measures how much different is A_i from the least optimal A^- .

Therefore, if we consider a map $f : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ that is decreasing in the first component and increasing on the second one, we obtain the following value a_i for alternative A_i :

$$a_i = f(d(A_i, A^+), d(A_i, A^-)).$$

Thus, the greater the value of a_i , the more preferred is the alternative A_i .

We can see that we can choose the function f depending on the part we are more interested in: the difference between A_i and the optimum A^+ or the difference between A_i and the least optimum A^- . The following examples illustrate this fact.

Example 5.79 ([211, Section 4]) *A city is planning to build a library, and the city commissioner has to determine the air-conditioning system to be installed in the library. The builder offers the commissioner five feasible alternatives A_i , which might be adapted to the physical structure of the library. Suppose that three attributes C_1 (economic), C_2 (functional) and C_3 (operational) are taken into consideration in the installation problem,*

and that the weight vector of the attributes C_j is $\alpha = (0.3, 0.5, 0.2)$. Assume moreover that the characteristics of the alternatives A_i are represented by the following IF-sets:

$$\begin{aligned} A_1 &= \{(C_1, 0.2, 0.4), (C_2, 0.7, 0.1), (C_3, 0.6, 0.3)\}, \\ A_2 &= \{(C_1, 0.4, 0.2), (C_2, 0.5, 0.2), (C_3, 0.8, 0.1)\}, \\ A_3 &= \{(C_1, 0.5, 0.4), (C_2, 0.6, 0.2), (C_3, 0.9, 0)\}, \\ A_4 &= \{(C_1, 0.3, 0.5), (C_2, 0.8, 0.1), (C_3, 0.7, 0.2)\}, \\ A_5 &= \{(C_1, 0.8, 0.2), (C_2, 0.7, 0), (C_3, 0.1, 0.6)\}. \end{aligned}$$

For these IF-sets, the corresponding A^+ and A^- are given by:

$$\begin{aligned} A^+ &= \{(C_1, 0.8, 0.2), (C_2, 0.8, 0), (C_3, 0.9, 0)\}. \\ A^- &= \{(C_1, 0.2, 0.5), (C_2, 0.5, 0.2), (C_3, 0.1, 0.6)\}. \end{aligned}$$

Then, if we consider the Hamming distance for IF-sets (see Subsection 5.1.3), we obtain the following:

	C_1	C_2	C_3
$l_{\text{IFS}}(A_1, A^+) - l_{\text{IFS}}(A_1 \cup \{C_j\}, A^+ \cup \{C_j\})$	1.2	0.2	0.6
$l_{\text{IFS}}(A_1, A^-) - l_{\text{IFS}}(A_1 \cup \{C_j\}, A^- \cup \{C_j\})$	0.2	0.4	1
$l_{\text{IFS}}(A_2, A^+) - l_{\text{IFS}}(A_2 \cup \{C_j\}, A^+ \cup \{C_j\})$	0.8	0.6	0.2
$l_{\text{IFS}}(A_2, A^-) - l_{\text{IFS}}(A_2 \cup \{C_j\}, A^- \cup \{C_j\})$	0.6	0	1.4
$l_{\text{IFS}}(A_3, A^+) - l_{\text{IFS}}(A_3 \cup \{C_j\}, A^+ \cup \{C_j\})$	0.6	0.4	0
$l_{\text{IFS}}(A_3, A^-) - l_{\text{IFS}}(A_3 \cup \{C_j\}, A^- \cup \{C_j\})$	0.6	0.2	1.6
$l_{\text{IFS}}(A_4, A^+) - l_{\text{IFS}}(A_4 \cup \{C_j\}, A^+ \cup \{C_j\})$	1	0.2	0.4
$l_{\text{IFS}}(A_4, A^-) - l_{\text{IFS}}(A_4 \cup \{C_j\}, A^- \cup \{C_j\})$	0.2	0.6	1.2
$l_{\text{IFS}}(A_5, A^+) - l_{\text{IFS}}(A_5 \cup \{C_j\}, A^+ \cup \{C_j\})$	0	0.2	1.6
$l_{\text{IFS}}(A_5, A^-) - l_{\text{IFS}}(A_5 \cup \{C_j\}, A^- \cup \{C_j\})$	1.2	0.4	0

Thus:

$$\begin{aligned} d(A_1, A^+) &= 0.3 \cdot 1.2 + 0.5 \cdot 0.2 + 0.2 \cdot 0.6 = 0.58. \\ d(A_1, A^-) &= 0.3 \cdot 0.2 + 0.5 \cdot 0.4 + 0.2 \cdot 1 = 0.46. \\ d(A_2, A^+) &= 0.3 \cdot 0.8 + 0.5 \cdot 0.6 + 0.2 \cdot 0.2 = 0.58. \\ d(A_2, A^-) &= 0.3 \cdot 0.6 + 0.5 \cdot 0 + 0.2 \cdot 1.4 = 0.46. \\ d(A_3, A^+) &= 0.3 \cdot 0.6 + 0.5 \cdot 0.4 + 0.2 \cdot 0 = 0.38. \\ d(A_3, A^-) &= 0.3 \cdot 0.6 + 0.5 \cdot 0.2 + 0.2 \cdot 1.6 = 0.6. \\ d(A_4, A^+) &= 0.3 \cdot 1 + 0.5 \cdot 0.2 + 0.2 \cdot 0.4 = 0.48. \\ d(A_4, A^-) &= 0.3 \cdot 0.2 + 0.5 \cdot 0.6 + 0.2 \cdot 1.2 = 0.6. \\ d(A_5, A^+) &= 0.3 \cdot 0 + 0.5 \cdot 0.2 + 0.2 \cdot 1.6 = 0.42. \\ d(A_5, A^-) &= 0.3 \cdot 1.2 + 0.5 \cdot 0.4 + 0.2 \cdot 0 = 0.56. \end{aligned}$$

Assume that we want to choose the alternative that is, at the same time, more similar to A^+ and less similar to the worst case A^- . In such a case we can consider the function f given by $f(x, y) = \frac{1}{2}(\frac{1}{x} + y)$. We can see that this function take into account the difference

between A_i and A^+ and between A_i and A^- . We obtain the following results:

$$\begin{aligned} a_1 &= f(d(A_1, A^+), d(A_1, A^-)) = \frac{1}{2} \left(\frac{1}{0.58} + 0.46 \right) = 1.09. \\ a_2 &= f(d(A_2, A^+), d(A_2, A^-)) = \frac{1}{2} \left(\frac{1}{0.58} + 0.46 \right) = 1.09. \\ a_3 &= f(d(A_3, A^+), d(A_3, A^-)) = \frac{1}{2} \left(\frac{1}{0.38} + 0.6 \right) = 1.62. \\ a_4 &= f(d(A_4, A^+), d(A_4, A^+)) = \frac{1}{2} \left(\frac{1}{0.48} + 0.6 \right) = 1.34. \\ a_5 &= f(d(A_5, A^+), d(A_5, A^+)) = \frac{1}{2} \left(\frac{1}{0.42} + 0.56 \right) = 1.47. \end{aligned}$$

Assume next that we decide to choose the alternative that is more similar to the optimum A^+ , regardless the difference from A^- . In that case, we may consider $f(x, y) = \frac{1}{x}$. This function only depends in the difference between A_i and the optimum A^+ . We obtain the following result:

$$\begin{aligned} a_1 &= f(d(A_1, A^+), d(A_1, A^-)) = \frac{1}{d(A_1, A^+)} = \frac{1}{0.58}. \\ a_2 &= f(d(A_2, A^+), d(A_2, A^-)) = \frac{1}{d(A_2, A^+)} = \frac{1}{0.58}. \\ a_3 &= f(d(A_3, A^+), d(A_3, A^-)) = \frac{1}{d(A_3, A^+)} = \frac{1}{0.38}. \\ a_4 &= f(d(A_4, A^+), d(A_4, A^+)) = \frac{1}{d(A_4, A^+)} = \frac{1}{0.48}. \\ a_5 &= f(d(A_5, A^+), d(A_5, A^+)) = \frac{1}{d(A_5, A^+)} = \frac{1}{0.42}. \end{aligned}$$

Thus, $A_3 \succ A_5 \succ A_4 \succ A_1 \sim A_2$, and as a consequence the best alternative is A_3 .

Finally, assume we are interested in the alternative that differs more from the worst alternative A^- . In such a situation we should consider $f(x, y) = y$. This function only depends on the difference between A_i and A^- . We obtain the following results:

$$\begin{aligned} a_1 &= f(d(A_1, A^+), d(A_1, A^-)) = d(A_1, A^-) = 0.46. \\ a_2 &= f(d(A_2, A^+), d(A_2, A^-)) = d(A_2, A^-) = 0.46. \\ a_3 &= f(d(A_3, A^+), d(A_3, A^-)) = d(A_3, A^-) = 0.6. \\ a_4 &= f(d(A_4, A^+), d(A_4, A^+)) = d(A_4, A^-) = 0.6. \\ a_5 &= f(d(A_5, A^+), d(A_5, A^+)) = d(A_5, A^-) = 0.56. \end{aligned}$$

Thus, $A_3 \sim A_4 \succ A_5 \succ A_1 \sim A_2$. We conclude that in this case A_3 and A_4 are the preferred alternatives. ♦

Example 5.80 Consider the previous example, but now with the Hausdorff distance for IF-sets (see Section 5.1.3). Using the same IVF-sets, we obtain that:

	C_1	C_2	C_3
$d_H(A_1, A^+) - d_H(A_1 \cup \{C_j\}, A^+ \cup \{C_j\})$	0.6	0.1	0.3
$d_H(A_1, A^-) - d_H(A_1 \cup \{C_j\}, A^- \cup \{C_j\})$	0.3	0.2	0.5
$d_H(A_2, A^+) - d_H(A_2 \cup \{C_j\}, A^+ \cup \{C_j\})$	0.4	0.3	0.1
$d_H(A_2, A^-) - d_H(A_2 \cup \{C_j\}, A^- \cup \{C_j\})$	0.3	0	0.7
$d_H(A_3, A^+) - d_H(A_3 \cup \{C_j\}, A^+ \cup \{C_j\})$	0.3	0.2	0
$d_H(A_3, A^-) - d_H(A_3 \cup \{C_j\}, A^- \cup \{C_j\})$	0.3	0.1	0.8
$d_H(A_4, A^+) - d_H(A_4 \cup \{C_j\}, A^+ \cup \{C_j\})$	0.5	0.1	0.2
$d_H(A_4, A^-) - d_H(A_4 \cup \{C_j\}, A^- \cup \{C_j\})$	0.3	0.3	0.6
$d_H(A_5, A^+) - d_H(A_5 \cup \{C_j\}, A^+ \cup \{C_j\})$	0	0.1	0.8
$d_H(A_5, A^-) - d_H(A_5 \cup \{C_j\}, A^- \cup \{C_j\})$	0.6	0.2	0

Then:

$$\begin{aligned}
d(A_1, A^+) &= 0.3 \cdot 0.6 + 0.5 \cdot 0.1 + 0.2 \cdot 0.3 = 0.29. \\
d(A_1, A^-) &= 0.3 \cdot 0.3 + 0.5 \cdot 0.2 + 0.3 \cdot 0.5 = 0.34. \\
d(A_2, A^+) &= 0.3 \cdot 0.4 + 0.5 \cdot 0.3 + 0.3 \cdot 0.1 = 0.3. \\
d(A_2, A^-) &= 0.3 \cdot 0.3 + 0.5 \cdot 0 + 0.3 \cdot 0.7 = 0.3. \\
d(A_3, A^+) &= 0.3 \cdot 0.3 + 0.5 \cdot 0.2 + 0.3 \cdot 0 = 0.19. \\
d(A_3, A^-) &= 0.3 \cdot 0.3 + 0.5 \cdot 0.1 + 0.3 \cdot 0.8 = 0.38. \\
d(A_4, A^+) &= 0.3 \cdot 0.5 + 0.5 \cdot 0.1 + 0.3 \cdot 0.2 = 0.26. \\
d(A_4, A^-) &= 0.3 \cdot 0.3 + 0.5 \cdot 0.3 + 0.3 \cdot 0.6 = 0.42. \\
d(A_5, A^+) &= 0.3 \cdot 0 + 0.5 \cdot 0.1 + 0.3 \cdot 0.8 = 0.29. \\
d(A_5, A^-) &= 0.3 \cdot 0.6 + 0.5 \cdot 0.2 + 0.3 \cdot 0 = 0.28.
\end{aligned}$$

As before, we first look for the alternative that is, at the same time, more similar to the optimum A^+ and less similar to the least optimum A^- . For this aim we can consider the function $f(x, y) = \frac{1}{2} \left(\frac{1}{x} + y \right)$. It holds that:

$$\begin{aligned}
a_1 &= f(d(A_1, A^+), d(A_1, A^-)) = \frac{1}{2} \left(\frac{1}{0.29} + 0.34 \right) = 3.79. \\
a_2 &= f(d(A_2, A^+), d(A_2, A^-)) = \frac{1}{2} \left(\frac{1}{0.3} + 0.3 \right) = 3.63. \\
a_3 &= f(d(A_3, A^+), d(A_3, A^-)) = \frac{1}{2} \left(\frac{1}{0.19} + 0.38 \right) = 5.64. \\
a_4 &= f(d(A_4, A^+), d(A_4, A^-)) = \frac{1}{2} \left(\frac{1}{0.26} + 0.42 \right) = 4.27. \\
a_5 &= f(d(A_5, A^+), d(A_5, A^-)) = \frac{1}{2} \left(\frac{1}{0.29} + 0.28 \right) = 3.72.
\end{aligned}$$

Then $A_3 \succ A_4 \succ A_1 \succ A_5 \succ A_2$, and therefore A_3 is the preferred alternative.

Next, we seek for the alternative that is more similar to the optimal A^+ . A possible

function f for this scenario is $f(x, y) = \frac{1}{x}$. In such a case:

$$\begin{aligned} a_1 &= f(d(A_1, A^+), d(A_1, A^-)) = \frac{1}{d(A_1, A^+)} = \frac{1}{0.29}. \\ a_2 &= f(d(A_2, A^+), d(A_2, A^-)) = \frac{1}{d(A_2, A^+)} = \frac{1}{0.3}. \\ a_3 &= f(d(A_3, A^+), d(A_3, A^-)) = \frac{1}{d(A_3, A^+)} = \frac{1}{0.19}. \\ a_4 &= f(d(A_4, A^+), d(A_4, A^+)) = \frac{1}{d(A_4, A^+)} = \frac{1}{0.26}. \\ a_5 &= f(d(A_5, A^+), d(A_5, A^+)) = \frac{1}{d(A_5, A^+)} = \frac{1}{0.29}. \end{aligned}$$

Then, it holds that $A_3 \succ A_4 \succ A_1 \sim A_5 \succ A_2$, and therefore alternative A_3 is the preferred one.

Finally, if we look for the alternative that differs more from the worst possibility A^- , we can choose $f(x, y) = y$. In that case,

$$\begin{aligned} a_1 &= f(d(A_1, A^+), d(A_1, A^-)) = d(A_1, A^-) = 0.34. \\ a_2 &= f(d(A_2, A^+), d(A_2, A^-)) = d(A_2, A^-) = 0.3. \\ a_3 &= f(d(A_3, A^+), d(A_3, A^-)) = d(A_3, A^-) = 0.38. \\ a_4 &= f(d(A_4, A^+), d(A_4, A^+)) = d(A_4, A^-) = 0.42. \\ a_5 &= f(d(A_5, A^+), d(A_5, A^+)) = d(A_5, A^-) = 0.28. \end{aligned}$$

We conclude that $A_4 \succ A_3 \succ A_1 \succ A_2 \succ A_5$, whence A_4 is the best alternative. ♦

5.3.3 Using IF-divergences to extend stochastic dominance

Consider now the problem of comparing more than two random variables. In Section 3.3 we mentioned that both stochastic dominance and statistical preference are methods for the pairwise comparison of random variables, and we proposed a generalization of statistical preference for comparing more than two random variables, based on an extension of the probabilistic relation defined in Equation (2.7). Now, based on the IF-divergences and due to the connection between IF-sets and imprecise probabilities we have investigated in Section 5.2, we propose a method that allows us to compare n p-boxes in order to obtain an order between them.

In order to do this, consider n p-boxes $(\underline{F}_1, \overline{F}_1), \dots, (\underline{F}_n, \overline{F}_n)$. For each p-box $(\underline{F}_i, \overline{F}_i)$, define the random interval Γ_i by $\Gamma_i(\omega) = [U_i(\omega), V_i(\omega)]$, where U_i and V_i are the quantile functions of \underline{F}_i and \overline{F}_i , respectively. Then, for each p-box $(\underline{F}_i, \overline{F}_i)$ we have an associated random interval that we can understand as a random interval defined from an IF-set A_i . Thus, we can apply the method described in Section 5.3.2 to obtain the p-box closer to the “optimal” p-box, that is the one associated with A^+ , and more distant to the “less optimal” p-box, that is the one associated with A^- .

Remark 5.81 *During this section we have investigated measures of comparison defined on finite spaces, according to the usual framework. However, all the measures we have studied can be extended to any space, non-necessarily finite. For instance, when dealing with local IF-divergences, they could be defined from $[a, b]$ to \mathbb{R} by using the Lebesgue measure $\lambda_{[a,b]}$ in $[a, b]$:*

$$D_{\text{IFS}}(A, B) = \int_{[a,b]} h_{\text{IFS}}(\mu_A(\omega), \nu_A(\omega), \mu_B(\omega), \nu_B(\omega)) d\lambda_{[a,b]}.$$

In order to illustrate this method, we propose a numerical example based on the comparison of sets of Lorenz Curves as we made in Section 4.4.1.

Numerical example: comparison of Lorenz curves

In Section 4.4.1 we considered the Lorenz curves associated with several countries. Such data was illustrated in Table 4.2, and Table 4.3 showed the cumulative distribution functions associated with each Lorenz curve. Recall that we grouped the countries by continents/regions in the following way:

- Group 1: China, Japan, India.
- Group 2: Finland, Norway, Sweden.
- Group 3: Canada, USA.
- Group 4: FYR Macedonia, Greece.
- Group 5: Australia, Maldives.

Next table shows the p-boxes associated with these groups.

Group		F(0.2)	F(0.4)	F(0.6)	F(0.8)	F(1)
Group-1	\overline{F}_1	47.81	69.81	84.47	94.27	100
	\underline{F}_1	35.65	57.63	75.21	89.42	100
Group-2	\overline{F}_2	37.23	59.33	76.9	90.88	100
	\underline{F}_2	36.63	58.84	76.31	90.38	100
Group-3	\overline{F}_3	45.82	68.22	83.88	94.56	100
	\underline{F}_3	39.94	62.89	80.07	92.8	100
Group-4	\overline{F}_4	41.49	64.53	81.37	93.26	100
	\underline{F}_4	37.43	60.04	77.53	90.98	100
Group-5	\overline{F}_5	49.24	66.9	82.61	94.1	100
	\underline{F}_5	41.32	64.89	82.09	93.49	100

Assume now that we are interested in comparing all the groups of countries together. Then, following the steps of Section 5.3.2, denote by A_i the IF-set defined by $\mu_{A_i} = \overline{F}_i^{-1}$ and $1 - \nu_{A_i} = \underline{F}_i^{-1}$, that is, the IF-set defined by the quantile functions of $(\underline{F}_i, \overline{F}_i)$. These IF-sets are given by:

$$\mu_{A_1}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 47.81]. \\ 0.4 & \text{if } t \in (47.81, 69.81]. \\ 0.6 & \text{if } t \in (69.81, 84.47]. \\ 0.8 & \text{if } t \in (84.47, 94.27]. \\ 1 & \text{if } t \in (94.27, 100]. \end{cases} \quad 1 - \nu_{A_1}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 35.64]. \\ 0.4 & \text{if } t \in (35.64, 57.63]. \\ 0.6 & \text{if } t \in (57.63, 75.21]. \\ 0.8 & \text{if } t \in (75.21, 89.42]. \\ 1 & \text{if } t \in (89.42, 100]. \end{cases}$$

$$\mu_{A_2}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 37.23]. \\ 0.4 & \text{if } t \in (37.23, 59.33]. \\ 0.6 & \text{if } t \in (59.33, 76.9]. \\ 0.8 & \text{if } t \in (76.9, 90.88]. \\ 1 & \text{if } t \in (90.88, 100]. \end{cases} \quad 1 - \nu_{A_2}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 36.63]. \\ 0.4 & \text{if } t \in (36.63, 58.84]. \\ 0.6 & \text{if } t \in (58.84, 76.31]. \\ 0.8 & \text{if } t \in (76.31, 90.38]. \\ 1 & \text{if } t \in (90.38, 100]. \end{cases}$$

$$\mu_{A_3}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 45.82]. \\ 0.4 & \text{if } t \in (45.82, 68.22]. \\ 0.6 & \text{if } t \in (68.22, 83.88]. \\ 0.8 & \text{if } t \in (83.88, 94.56]. \\ 1 & \text{if } t \in (94.56, 100]. \end{cases} \quad 1 - \nu_{A_3}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 39.94]. \\ 0.4 & \text{if } t \in (39.94, 62.89]. \\ 0.6 & \text{if } t \in (62.89, 80.07]. \\ 0.8 & \text{if } t \in (80.07, 92.8]. \\ 1 & \text{if } t \in (92.8, 100]. \end{cases}$$

$$\mu_{A_4}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 41.49]. \\ 0.4 & \text{if } t \in (41.49, 64.53]. \\ 0.6 & \text{if } t \in (64.53, 81.37]. \\ 0.8 & \text{if } t \in (81.37, 93.26]. \\ 1 & \text{if } t \in (93.26, 100]. \end{cases} \quad 1 - \nu_{A_4}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 37.43]. \\ 0.4 & \text{if } t \in (37.43, 60.04]. \\ 0.6 & \text{if } t \in (60.04, 77.53]. \\ 0.8 & \text{if } t \in (77.53, 90.98]. \\ 1 & \text{if } t \in (90.98, 100]. \end{cases}$$

$$\mu_{A_5}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 49.24]. \\ 0.4 & \text{if } t \in (49.24, 66.9]. \\ 0.6 & \text{if } t \in (66.9, 82.61]. \\ 0.8 & \text{if } t \in (82.61, 94.1]. \\ 1 & \text{if } t \in (94.1, 100]. \end{cases} \quad 1 - \nu_{A_5}(t) = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 41.32]. \\ 0.4 & \text{if } t \in (41.32, 64.89]. \\ 0.6 & \text{if } t \in (64.89, 82.09]. \\ 0.8 & \text{if } t \in (82.09, 93.49]. \\ 1 & \text{if } t \in (93.49, 100]. \end{cases}$$

Consider now the IF-sets A^+ and A^- defined in Equations (5.11) and (5.12), that are defined by $\mu_{A^+} = \mu_{A_2}$, $1 - \nu_{A^+} = 1 - \nu_{A_1}$, $1 - \nu_{A^-} = 1 - \nu_{A_5}$ and:

$$\mu_{A^-} = \begin{cases} 0 & \text{if } t = 0. \\ 0.2 & \text{if } t \in (0, 49.24]. \\ 0.4 & \text{if } t \in (49.24, 69.81]. \\ 0.6 & \text{if } t \in (69.81, 84.47]. \\ 0.8 & \text{if } t \in (84.47, 94.56]. \\ 1 & \text{if } t \in (94.56, 100]. \end{cases}$$

Now, we consider two of the most usual measures of comparison of IF-divergences we can find in the literature, the Hausdorff and the Hamming distances that, as we have said in Section 5.1.3, are also local IF-divergences. Recall that they are defined, respectively, by:

$$d_H(A, B) = \int_0^{100} \max\{|\mu_A(\omega) - \mu_B(\omega)|, |\nu_A(\omega) - \nu_B(\omega)|\}d\omega.$$

$$l_{IFS}(A, B) = \frac{1}{2} \int_0^{100} |\mu_A(\omega) - \mu_B(\omega)| + |\nu_A(\omega) - \nu_B(\omega)| + |\pi_A(\omega) - \pi_B(\omega)|d\omega.$$

We represent the results on the next table.

	$l_{IFS}(A_i, A^+)$	$l_{IFS}(A_i, A^-)$	$d_H(A_i, A^+)$	$d_H(A_i, A^-)$
A_1	6.404	2.561	6.404	5.122
A_2	0.852	4.773	0.852	7.414
A_3	4.594	1.169	6.916	5.974
A_4	2.439	3.324	5.052	6.386
A_5	5.448	0.523	6.99	1.046

Now, we consider three different functions:

$$f_1(x, y) = y, \quad f_2(x, y) = -x \quad \text{and} \quad f_3(x, y) = y - x.$$

f_1 only focus in the closest IF-set to the least optimal alternative; f_2 only focus in the closest IF-set to the most optimal alternative, while f_3 focus in the IF-set that is both closer to the most optimal alternative and less closer IF-set to the least optimal alternative. We obtain the following results:

l_{IFS}	f_1	f_2	f_3	d_{H}	f_1	f_2	f_3
A_1	2.561	-6.404	-3.843	A_1	5'122	-6'404	1'282
A_2	4.773	-0.852	3.921	A_2	7'414	-0'852	6'562
A_3	1.169	-4.594	-3.425	A_3	5'974	-6'916	-0'942
A_4	3.324	-2.439	0.885	A_4	6'386	-5'052	1'334
A_5	0.523	-5.448	-4.925	A_5	1'046	-6'99	-5'944

In the three cases, and with both IF-divergences, the preferred group is the second, that is, the group of Nordic countries. The worst alternative, except for the IF-divergence l_{IFS} and the function f_2 , is the group A_5 , that is the group of oceanic countries. This means that the group of countries that has a better wealth distribution is the group of Nordic countries, while the greater wealth inequalities are, in the most cases, in the group of oceanic countries.

5.4 Conclusions

The comparison of fuzzy sets is a topic that has been widely investigated, an several papers with mathematical theories can be found in the literature. However, when we move towards IF-sets the efforts are somewhat scattered, and there is not an axiomatic approach to the comparison of this kind of sets.

For this reason we have developed a mathematical theory of the comparison of IF-sets. In particular, we have focused on IF-divergences, which are more restrictive measures than IF-dissimilarities. In particular, IF-divergences with the local property, named local IF-divergences, played an important role. As was expected, a connection between divergences for fuzzy sets and IF-divergences can be established, and we have found the conditions under which the local property, among other interesting properties, are preserved when we move from IF-divergences to divergences, and conversely, from divergences to IF-divergences. We also showed that these measures can be applied in pattern recognition and decision making, showing several examples.

On the other hand, we have investigated the connection between IVF-sets and Imprecise Probabilities. In this sense, we assumed that the IVF-set is defined on a probability space, and then it can be interpreted as a random interval. Then, we have

investigated the probabilistic information encoded by the random interval or its measurable selections, and we found conditions under which this probabilistic information coincides with the probabilistic information given by its associated set of probabilities dominated by the upper probability. We also investigated the connection between our approach and other ones that can be found in the literature. In particular, the definition of probability for IF-sets given by Grzegorzewski and Mrowka is contained as a particular case of our theory.

The connection between IVF-sets and Imprecise Probabilities has allowed us to extend stochastic dominance to the comparison of more than two p-boxes simultaneously, determining also a complete relationship (i.e., avoiding incomparability). This method, that depends on the chosen IF-divergence, gives us a ranking of the p-boxes. We have illustrated its behaviour continuing with the example of Section 4.4.1 in which we compare sets of Lorenz Curves.

For future research, some open problems arise in the topic of comparison of IF-sets. On the one hand, it is possible to investigate under which conditions IF-divergences, and in particular local IF-divergences, can define an entropy for IF-sets ([29]). On the other hand, as could be seen in the applications of IF-divergence, it is interesting to introduce weights in the elements of the universe. In this situation it would be interesting to define local IF-divergence with weights, and trying to find an analogous result to Theorem 5.29 to characterize them. Furthermore, we could investigate if it is possible to define locality with an operator different than the sum; a t-conorm for instance. Moreover, our aim is to extend the local property to general universes, non-necessarily finite. With respect to the connection between IF-sets, IVF-sets and Imprecise Probabilities, we pretend to continue studying IF-sets and IVF-sets as bipolar models for representing positive and negative information ([72, 73]).

Conclusiones y trabajo futuro

A lo largo de esta memoria se ha tratado el problema de la comparación de alternativas bajo ciertos tipos de falta de información: incertidumbre e imprecisión. La incertidumbre se refiere a situaciones en las que los posibles resultados del experimento están perfectamente descritos, pero el resultado del mismo es desconocido. Por otra parte, la imprecisión se refiere a situaciones en las que el resultado del experimento es conocido pero no es posible describirlo con precisión. Las herramientas utilizadas para modelar la incertidumbre y la imprecisión han sido la Teoría de las Probabilidades y la Teoría de los Conjuntos Intuicionísticos, respectivamente, mientras que la Teoría de las Probabilidades Imprecisas se ha utilizado para modelar ambas faltas de información simultáneas.

Cuando las alternativas a comparar están definidas bajo incertidumbre, éstas se han modelado mediante variables aleatorias, que son habitualmente comparadas mediante órdenes estocásticos. En esta memoria se han considerado, principalmente, dos de estos órdenes: la dominancia estocástica y la preferencia estadística. El primero de ellos es el orden estocástico más habitual en la literatura, y ha sido utilizado en diferentes ámbitos con destacables resultados. Por otra parte, la preferencia estadística es el método más adecuado para comparar variables cualitativas.

A pesar de que la dominancia estocástica es un método que ha sido investigado por varios autores, la preferencia estadística no ha sido estudiada con tanta profundidad. Ésta es la razón por la cual hemos estudiado sus propiedades como orden estocásticos. Uno de los resultados más destacados en este estudio es la relación de este método con la mediana. Esto demuestra que, mientras que la dominancia estocástica está relacionada con la media, la preferencia estadística es más cercana a otro parámetro de localización.

También hemos investigado la relación entre la dominancia estocástica y la preferencia estadística, y hemos encontrado condiciones bajo las cuales la dominancia estocástica de primer orden implica la preferencia estadística. Dado que la preferencia estadística depende de la distribución conjunta de las variables y , por tanto, de la cópula que las liga, dichas condiciones están también relacionadas con la cópula. El Teorema 3.64 resume estas condiciones: variables aleatorias independientes, variables aleatorias continuas ligadas por una cópula Arquimediana o variables aleatorias o bien continuas o bien discretas

con soportes finitos que son comonótonas o contramonótonas. Además, hemos comprobado que esta relación no se cumple en general. Por tanto, de manera natural surge la siguiente cuestión: ¿es posible caracterizar las cópulas que hacen que la dominancia estocástica de primer orden implique la preferencia estadística?

Cuando las variables a comparar pertenecen a la misma familia paramétrica de distribuciones, como por ejemplo Bernoulli, exponencial, uniforme, Pareto, beta o normal, hemos visto que la dominancia estocástica y la preferencia estadística coinciden, y de hecho, ambos métodos se reducen a la comparación de sus esperanzas. Por esta razón es posible plantearse la siguiente conjetura: cuando las variables a comparar siguen la misma distribución perteneciente a la familia exponencial de distribuciones, tanto la dominancia estocástica como la preferencia estadística se reducen a la comparación de esperanzas y son, por tanto, equivalentes. Aunque éste es un problema abierto, una primera aproximación basada en simulaciones se ha realizado en [32].

La dominancia estocástica y la preferencia estadística son métodos de comparación de variables aleatorias por pares. Esto hace que en ocasiones no sean métodos adecuados para comparar más de dos variables simultáneamente. De hecho, la preferencia estadística es una relación no transitiva, y por lo tanto puede producir resultados ilógicos. Ésta es la razón que nos ha llevado a definir una generalización de la preferencia estadística para la comparación de más de dos variables simultáneamente. Siguiendo la misma aproximación que en el caso de la preferencia estadística, nuestra generalización da un grado de preferencia a cada una de las variables de manera que todos los grados sumen uno. Por lo tanto, la variable preferida será aquella con el mayor grado de preferencia. Para este método hemos estudiado su conexión con los órdenes estocásticos por pares. En particular, hemos visto que las mismas condiciones del Teorema 3.64 permiten asegurar que si una de las variables domina estocásticamente de primer grado al resto, entonces ésta es también preferida a todas las demás utilizando nuestra generalización de la preferencia estadística.

A la preferencia estadística general le podemos dar la siguiente interpretación: dado un conjunto de alternativas (en este caso variables aleatorias) tenemos que elegir entre la preferida, y podemos asignar a cada variable un grado de preferencia. Este grado de preferencia puede entenderse como cuánto de preferida es cada alternativa sobre el resto. Esto hace que la preferencia estadística general se pueda ver como una función de elección difusa ([81, 207]). Un punto abierto sería por tanto estudiar la preferencia estadística general como una función de elección difusa.

Hay situaciones en las cuales las alternativas a comparar están definidas tanto bajo incertidumbre como bajo imprecisión. En tales casos, las variables aleatorias no recogen todas la información. En esta situación hemos modelado las alternativas mediante conjuntos de variables aleatorias con una interpretación epistémica: cada conjunto contiene la variable aleatoria original, que es desconocida. De cara a comparar estos conjuntos de alternativas, hemos tenido que extender los órdenes estocásticos para la comparación de conjuntos de variables aleatorias. Esta extensión da lugar a seis posibles métodos de

ordenación de conjuntos de variables aleatorias. Una vez investigadas estas extensiones, nos hemos centrado en los casos en los que el orden estocástico utilizado es o bien la dominancia estocástica o bien la preferencia estadística, y hemos llamado a sus extensiones dominancia estocástica imprecisa y preferencia estadística imprecisa. La Proposición 4.19 y el Corolario 4.22 muestran que la dominancia estocástica imprecisa está relacionada con la comparación de las p -boxes asociadas a los conjuntos de variables aleatorias por medio de la dominancia estocástica. Estos resultados también nos permiten ver el estudio realizado por Denoeux ([61]) como un caso particular de nuestro estudio. Denoeux consideró dos medidas de creencia, y sus medidas de plausibilidad asociadas, y utilizó la dominancia estocástica para compararlas. Sin embargo, dado que las medidas de creencia y plausibilidad definen conjuntos de probabilidades, es posible compararlas mediante la dominancia estocástica imprecisa.

Lo mismo ocurre con posibilidades: una medida de posibilidad define un conjunto de probabilidades, y por lo tanto es posible utilizar la dominancia estocástica imprecisa para compararlas. En la Proposición 4.52 hemos dado una caracterización de la dominancia estocástica imprecisa para medidas de posibilidad con distribuciones de posibilidad continuas. Aquí surge un nuevo problema abierto: en caso de que las distribuciones de posibilidad asociadas a las distribuciones de posibilidad no sean continuas, ¿se cumple la misma caracterización de la Proposición 4.52?

Dos situaciones habituales dentro de la Teoría de la Decisión se pueden modelar mediante la comparación de conjuntos de variables aleatorias. Por una parte, hemos considerado la comparación de dos variables aleatorias con imprecisión en las utilidades. Esta falta de información ha sido modelada con conjuntos aleatorios. La información probabilística de un conjunto aleatorio se recoge en sus selecciones medibles. Por tanto, la comparación de conjuntos aleatorios se realiza mediante la comparación de sus conjuntos de selecciones medibles. Por otra parte, hemos considerado la comparación de variables aleatorias definidas sobre un espacio probabilístico donde la probabilidad no está definida de manera precisa. En esta situación, en vez de haber una única probabilidad, hemos considerado un conjunto de probabilidades. De esta manera también es posible definir dos conjuntos de variables aleatorias que recogen la información disponible. Para estas dos situaciones hemos investigado en particular las propiedades de la dominancia estocástica imprecisa y la preferencia estadística imprecisa, estudiando sus conexiones con la Teoría de las Probabilidades Imprecisas.

La preferencia estadística es un orden estocástico que está basado en la distribución conjunta de las variables aleatorias. El Teorema de Sklar asegura que la función de distribución conjunta de dos variables se puede expresar a través de las marginales mediante el uso de la cópula adecuada. Ahora bien, dados dos variables aleatorias definidas en un espacio de probabilidad descrito de manera imprecisa, el Teorema de Sklar no permite construir la distribución conjunta. Para tratar este problema, hemos investigado las p -boxes bivariantes y su conexión con las probabilidades inferiores coherentes. En particular, hemos visto que las funciones de distribución inferior y superior asociadas

a un conjunto de funciones de distribución bivariantes no son en general funciones de distribución bivariantes, puesto que no cumplen la desigualdad de los rectángulos. Sin embargo, hemos visto que permiten definir una probabilidad inferior coherente, y a partir de resultados conocidos, las funciones de distribución inferior y superior cumplen cuatro desigualdades, llamadas (I-RI1), (I-RI2), (I-RI3) y (I-RI4), que pueden verse como las versiones imprecisas de la desigualdad de los rectángulos. La Proposición 4.102 asegura que dos funciones de distribución bivariantes, normalizadas y ordenadas definen una probabilidad inferior coherente cuando una de las funciones de distribución está definida sobre un espacio binario. Como trabajo futuro, deseamos estudiar si esta propiedad se cumple para funciones de distribución definidas sobre todo tipo de espacios, no necesariamente binarios.

El estudio de las p-boxes bivariantes nos han permitido demostrar una versión imprecisa del Teorema de Sklar. En nuestro estudio hemos asumido que partimos de dos distribuciones marginales imprecisas, definidas mediante p-boxes, y de un conjunto de cópulas. En esta situación es posible definir una p-box bivalente que defina a su vez una probabilidad inferior coherente. Además, hemos visto que el recíproco no se cumple en general, puesto que una p-box bivalente que define una probabilidad inferior coherente no puede ser expresada, en general, a través de las p-boxes marginales. Hemos comprobado que esta versión imprecisa del Teorema de Sklar es muy útil cuando hay que utilizar órdenes estocásticos bajo imprecisión.

La extensión de los órdenes estocásticos para la comparación de conjuntos de variables aleatorias tiene varias aplicaciones. Además de las aplicaciones habituales de los órdenes estocásticos en la Teoría de la Decisión, hemos visto que también pueden ser aplicados a la comparación de Curvas de Lorenz asociadas a distintos grupos de países o regiones. Estos conjuntos de Curvas de Lorenz han sido comparados mediante la dominancia estocástica imprecisa. Un estudio similar se ha realizado para comparar tasas de supervivencia asociadas a distintos tipos cáncer, estudiando qué tipo de cáncer tiene peor diagnóstico.

Las alternativas definidas bajo imprecisión, sin incertidumbre, se han modelado mediante conjuntos intuicionísticos (IF-sets). IF-sets son un tipo de conjuntos que sirven para modelar información bipolar: considera los grados de pertenencia y no pertenencia. Varios ejemplos de medidas de comparación de IF-sets se pueden encontrar en la literatura. Sin embargo, hasta este momento no se había desarrollado una teoría matemática. Por esta razón hemos considerados diferentes tipos de medidas de comparación, IF-disimilaridades, IF-divergencias, IF-disimilitudes y distancias, y las hemos estudiado desde un punto de vista teórico. Por una parte hemos estudiado las relaciones existentes entre estas medidas, y hemos definido una medida general de comparación de IF-sets que contiene a las otras medidas como casos particulares. Posteriormente nos hemos centrado en el estudio de las IF-divergencias, estudiando sus propiedades más interesantes. En particular, hemos considerado una clase de IF-divergencias que satisface una condición de localidad. También hemos visto qué conexión existe entre las divergen-

cias para conjuntos difusos y las IF-divergencias. Por último, se han explicado posibles aplicaciones de las IF-divergencias en el reconocimiento de patrones y en la Teoría de la Decisión.

Pasamos a comentar algunos problemas abiertos relacionados con las IF-divergencias. Por una parte, en caso de que los elementos del espacio inicial tengan unos pesos asociados, parece posible extender las IF-divergencias locales considerando los pesos. Por otra parte, las IF-divergencias se podrían estudiar como entropías para IF-sets. Además, creemos que es posible extender la propiedad de localidad para universos no finitos, o incluso dar una definición de localidad basada en un operador diferente de la suma, como podría ser una t-conorma.

En las últimas fechas varios investigadores han centrado su atención en cómo las probabilidades imprecisas pueden modelar la información bipolar. Dado que los IF-sets también son utilizados en este mismo contexto, hemos establecido una conexión entre ambas teorías. Para ello, hemos considerado IF-sets definidos en un espacio probabilístico, y si entendemos los IF-sets como conjuntos intervalo-valorados, pueden ser vistos como conjuntos aleatorios. En esta situación, la información probabilística está recogida en el conjunto de selecciones medibles. Hemos visto condiciones bajo las cuales esta información coincide con la información probabilística dada por el conjunto credal asociado al conjunto aleatorio. Además, hemos visto que aproximaciones que ya se encontraban en la literatura se pueden ver como casos particulares de nuestro estudio.

La conexión entre los IF-sets y las probabilidades imprecisas nos han permitido extender la dominancia estocástica para la comparación de más de dos p-boxes al mismo tiempo. Como trabajo futuro, pensamos que este estudio podría ser completado. En particular, se podría estudiar la relación del procedimiento que hemos explicado con el uso de la habitual distancia de Kolmogorov entre funciones de distribución. Sin embargo, creemos que éste puede verse como un caso particular de nuestro estudio.

Conclusions and further research

This memory has dealt with the problem of comparing alternatives under lack of information. This lack of information can be of different kinds, and here we have assumed that it corresponds to either uncertainty or imprecision. Uncertainty refers to situations where the possible results of the experiment are precisely described, but the exact result is unknown; on the other hand, imprecision refers to situations in which the result of the experiment is known but it cannot be precisely described. In order to model uncertainty and imprecision we have used Probability Theory and Intuitionistic Fuzzy Set Theory, respectively; when both these features appear together in the decision problem, we have used the Theory of Imprecise Probabilities.

When the alternatives are subject to uncertainty in the outcomes, we we have modelled them as random variables, and have used stochastic orders so as to make a comparison between them. We have focused mainly in two different stochastic orders: stochastic dominance and statistical preference. The former is one of the most widely used stochastic orders we can find in the literature and the latter is of particular interest when comparing qualitative variables. Indeed, although stochastic dominance is a well-known method that has been widely investigated by several authors, statistical preference remained partly unexplored. For this reason we have studied several properties of this stochastic order. Possibly the most important one is its characterization in terms of the median, that serves us to compare it as a robust alternative to stochastic dominance, which is related to another location parameter: the mean.

We have also investigated the relationship between stochastic dominance and statistical preference, and we have found conditions under which (first degree) stochastic dominance implies statistical preference. Since statistical preference depends on the copula that links the variables into a joint distribution, the conditions we have obtained are also related to the copula. Theorem 3.64 summarizes such conditions: independent random variables, continuous random variables coupled by an Archimedean copula and either continuous or discrete random variables with finite supports that are either comonotonic or countermonotonic. In addition, we have also showed that the implication between these two stochastic orders does not hold in general. Thus, the first open question naturally arises: it is possible to characterize the set of copulas that makes first

degree stochastic dominance to imply statistical preference?

When the random variables to be compared belong to the same parametric family of distributions, like for instance Bernoulli, exponential, uniform, Pareto, beta or normal, we have seen that both stochastic dominance and statistical preference coincide, and in fact, they are equivalent to compare the expectations of the random variables. This makes us to conjecture that when comparing two random variables that belong to the same parametric family of distribution within the exponential family, then stochastic dominance and statistical preference reduce to the comparison of the expectations. Although this problem is still open, a first approach, based on simulations, has already been done in [32].

Stochastic dominance and statistical preference are pairwise methods of comparison of random variables. In this respect, they were not defined to compare more than two variables simultaneously. In fact, statistical preference is not a transitive relation, and therefore it may produce nonsensical results. For this reason we have generalized statistical preference to the comparison of more than two random variables at the same time. With similar underlying ideas to those of statistical preference, our generalization assigns a preference degree to any of the random variables, and the sum of these preference degrees is one. Then, the preferred random variable is the one with greater preference degree. For this new approach we have investigated its connection to the usual statistical preference and stochastic dominance. In fact, the same conditions of Theorem 3.64 that guarantee that stochastic dominance implies statistical preference also assures that if there is a random variable that pairwise dominates all the others with respect to stochastic dominance, then such random variable will be the preferred one with respect to our generalization of statistical preference.

A future line of research appears associated with this general statistical preference. Given a set of alternatives (in this case, random variables) out of which we have to choose the preferred one, we can assign a degree of preference, that we understand as the strength of the preference of each alternative over the other. Then, the general statistical preference can be seen as a fuzzy choice function defined on a set of alternatives ([81, 207]). Thus, it may be interesting to investigate the properties of the general statistical preference in the framework of fuzzy choice functions.

On the other hand, there are situations in which the alternatives to be compared are defined, not only under uncertainty, but also under imprecision. In such cases, random variables do not collect all the available information. Thus, we have modelled the alternatives by means of sets of random variables with an epistemic interpretation: each set contains the real unknown random variable. In order to compare these sets, we need to extend stochastic orders to this general framework. In order to do this, we have considered any binary relation defined for the comparison of single random variables and we have extended it for the comparison of sets of random variables. We have thus considered six possible ways of ordering sets of random variables. After investigating some general properties of these extensions, we have focused in the cases in which binary relation is

either stochastic dominance or statistical preference. We have called their extensions imprecise stochastic dominance and imprecise statistical preference. Proposition 4.19 and Corollary 4.22 showed that the former is clearly connected to the comparison of the bounds of the associated p-boxes by means of stochastic dominance. These results also helped to show that the approach given by Denoeux ([61]) is a particular case of our more general framework. Denoeux considered two belief functions, and their respective plausibility functions, and used stochastic dominance to compare them. Since each belief and plausibility function can be represented as a set of probabilities, and therefore imprecise stochastic dominance can be applied; we have seen that our definitions become the ones given by Denoeux for this particular case.

The same happens with possibilities: each possibility defines a set of probabilities, and therefore the imprecise stochastic dominance can be used to compare them. Proposition 4.52 showed a characterization of the imprecise stochastic dominance for possibility measures with continuous possibility distribution. Thus, an open problem is to investigate if such characterization also holds for possibility measures with non-continuous possibility distributions.

We have explored two situations that are usually present in decision making and that can be modelled by means of the comparison of sets of random variables. On the one hand, we have considered the comparison of two random variables with imprecision on the utilities. We have modelled this imprecision with random sets. Since under our epistemic interpretation the set of measurable selections of a random set encodes its probabilistic information, the comparison of random sets must be made by means of the comparison of their associated credal sets. On the other hand, we can also compare random variables defined on a probability space with a non-precisely determined probability; in that case, we have to consider a set of probabilities instead a single one. In this situation we can also consider two sets of random variables that summarise all the available information. For these two particular situations we have explored the properties of imprecise stochastic dominance and statistical preference, and we have investigated their connection to imprecise probabilities.

We know that statistical preference is a stochastic order that is based on the joint distribution of the random variables. By Sklar's Theorem, this joint distribution is determined combining the marginals by means of a copula. However, given two random variables defined in a probability space with imprecise beliefs, Sklar's Theorem does not allow to define the joint distribution. In order to solve this problem, we have investigated bivariate p-boxes and how they can define a coherent lower probability. In particular, we have seen that the lower and upper distributions associated with a set of bivariate distribution functions are not in general bivariate distribution functions because they violate the rectangle inequality. However, we have seen that they define a coherent lower probability and they satisfy four inequalities, named (I-RI1), (I-RI2), (I-RI3) and (I-RI4), that can be seen as the imprecise versions of the rectangle inequality. We have seen in Proposition 4.102 that given two ordered normalized bivariate distribution functions that

satisfy them, they define a coherent lower probability if one of the normalized functions is defined on a binary space. An open problem for future research is to investigate if this property also holds for normalized functions defined on any space.

The study of bivariate p-boxes have allowed to define an imprecise version of Sklar's Theorem. We have assumed that we have two imprecise marginal distributions, that we model by means of p-boxes, and we have a set of possible copulas that link them. In this situation it is possible to define a bivariate p-box that defines a coherent lower probability. However, the second part of the Sklar's Theorem does not hold, because a bivariate p-box that defines a coherent lower probability cannot be expressed, in general, by means of the marginal p-boxes. We have also seen how this imprecise version is very useful when dealing with bivariate stochastic orders with imprecision.

The extension of stochastic orders to the comparison of sets of random variables we have proposed has several applications. Besides the usual application of stochastic orders in decision making, we have seen that they can be also applied to the comparison of the inequality indices between groups of countries. In this work, we have considered the Lorenz curve of each country, that measures the inequality of such country, and we have grouped them by geographical areas. Then, we have compared these groups of Lorenz curves using the imprecise stochastic dominance. We have made a similar approach to the comparison of cancer survival rates, grouping them by cancer sites, and we have analyzed which cancer site has a worst prognosis.

Alternatives defined under imprecision, without uncertainty, have been modelled by means of IF-sets. IF-sets are bipolar models that allow to define membership and non-membership degrees. Several examples of measures of comparison of IF-sets had been proposed in the literature. However, a mathematical theory had not been developed. For this reason we have considered different kinds of measures, IF-dissimilarities, IF-divergences, IF-dissimilitudes and distances, and we have investigated them from a theoretical point of view. First of all, we have seen the relationships between these measures, and we have defined a general measure of comparison of IF-sets that contains them as particular cases. Then, we have focused on IF-divergences and we have investigated its main properties. In particular, we have considered one instance of IF-divergences, those that satisfy a local property. We have also seen the connection between IF-divergences and divergences for fuzzy sets. We have also showed how IF-divergences can be applied within pattern recognition and decision making.

There are several open problems related to this study of IF-divergences. On the one hand, it would be interesting to define local IF-divergences that take into account a weight function on the the elements of the initial space. On the other hand, IF-divergences could be studied as entropies for IF-sets. Furthermore, it is possible to extend the local property to spaces non-necessarily finite, and also to define the local property by means of an operator different than the sum, like t-conorms, for instance.

Currently, several authors have been investigating how imprecise probabilities can

be used to model bipolar information. Since IF-sets are also useful in this context, we have established a connection between both theories. We have assumed that IF-sets are defined in a probability space; if we understand them as IVF-sets, they can then be seen as random sets. In that case, their probabilistic information can be encoded by the set of measurable selections. We have seen conditions under which such information coincides with the probabilistic information given the credal set associated to the random set. Furthermore, we have seen how previous approaches made for defining a probability measure on IF-sets can be embedded into our approach.

The connection between IF-sets and imprecise probabilities has allowed us to extend stochastic dominance to the comparison of more than two p-boxes simultaneously. For future research, we think that this proposal could be studied more thoroughly. For instance, a similar extension of stochastic dominance may be made by using the usual Kolmogorov distance between cumulative distribution functions. It would be interesting to determine if this becomes a particular case of our more general framework.

A Appendix: Basic Results

In this Appendix we prove some results that we have used throughout this report.

Lemma A.1 *Let a, b and c be three real numbers in $[0, 1]$. Then*

- a) $\max\{0, \min\{a, c\} - \min\{b, c\}\} \leq \max\{0, a - b\}$ and
 $\max\{0, \max\{a, c\} - \max\{b, c\}\} \leq \max\{0, a - b\}$.
- b) $\max(|\max\{a, c\} - \max\{b, c\}|, |\min\{a, c\} - \min\{b, c\}|) \leq |a - b|$.

Proof: We distinguish the following cases, depending on the minimum and the maximum of $\{a, c\}$ and $\{b, c\}$:

1. Assume that $\min\{a, c\} = a$ and $\min\{b, c\} = b$, and consequently $\max\{a, c\} = \max\{b, c\} = c$. Then:

- a) $\max\{0, \min\{a, c\} - \min\{b, c\}\} = \max\{0, a - b\}$.
 $\max\{0, \max\{a, c\} - \max\{b, c\}\} = 0 \leq \max\{0, a - b\}$.
- b) $|\max\{a, c\} - \max\{b, c\}| = |c - c| = 0 \leq |a - b|$.
 $|\min\{a, c\} - \min\{b, c\}| = |a - b|$.

2. Assume next that $\min\{a, c\} = a$ and $\min\{b, c\} = c$, and therefore $\max\{a, c\} = c$ and $\max\{b, c\} = b$. Note that, since $\min\{a, c\} = a$, then $a \leq c$, and therefore $a - c \leq 0$. Moreover, it also holds that $c \leq b$, and consequently $a \leq c \leq b$. Hence:

- a) $\max\{0, \min\{a, c\} - \min\{b, c\}\} = \max\{0, a - c\} = 0$
 $\leq \max\{0, a - b\}$.
 $\max\{0, \max\{a, c\} - \max\{b, c\}\} = \max\{0, c - b\} = 0$
 $\leq \max\{0, c - b\}$.
- b) $|\max\{a, c\} - \max\{b, c\}| = |c - b| \leq |a - b|$.
 $|\min\{a, c\} - \min\{b, c\}| = |a - c| \leq |a - b|$.

3. Thirdly, assume that $\min\{a, c\} = c$ and $\min\{b, c\} = b$, whence $\max\{a, c\} = a$ and $\max\{b, c\} = c$. In such a case, $c \leq a$ and $b \leq c$, and therefore $b \leq c \leq a$, that implies $c - b \leq a - b$ and $a - c \leq a - b$. Hence:

$$\begin{aligned} a) \quad & \max\{0, \min\{a, c\} - \min\{b, c\}\} = \max\{0, c - b\} \\ & \leq \max\{0, a - b\}. \\ & \max\{0, \max\{a, c\} - \max\{b, c\}\} = \max\{0, a - c\} \\ & \leq \max\{0, a - b\}. \\ b) \quad & |\max\{a, c\} - \max\{b, c\}| = |a - c| \leq |a - b|. \\ & |\min\{a, c\} - \min\{b, c\}| = |c - b| \leq |a - b|. \end{aligned}$$

4. Finally, assume that $\min\{a, c\} = \min\{b, c\} = c$, and consequently $\max\{a, c\} = a$ and $\max\{b, c\} = b$. Then:

$$\begin{aligned} a) \quad & \max\{0, \min\{a, c\} - \min\{b, c\}\} = 0 \leq \max\{0, a - b\}. \\ & \max\{0, \max\{a, c\} - \max\{b, c\}\} = \max\{0, a - b\}. \\ b) \quad & |\max\{a, c\} - \max\{b, c\}| = |a - b|. \\ & |\min\{a, c\} - \min\{b, c\}| = |c - c| = 0 \leq |a - b|. \quad \blacksquare \end{aligned}$$

Lemma A.2 *If (a_1, a_2) , (b_1, b_2) and (c_1, c_2) are elements on $\mathcal{T} = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}$, it holds that:*

$$\begin{aligned} \alpha &= |a_1 - b_1| + |a_2 - b_2| + |a_1 + a_2 - b_1 - b_2| \\ &\geq |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\ &\quad + |\max\{a_1, c_1\} + \min\{a_2, c_2\} - \max\{b_1, c_1\} - \min\{b_2, c_2\}| = \beta. \end{aligned}$$

Proof: Let us consider the following possibilities:

1. $a_1, b_1 \leq c_1$ and $a_2, b_2 \leq c_2$. Then:

$$\begin{aligned} \beta &= |c_1 - c_1| + |a_2 - b_2| + |c_1 + a_2 - c_1 - b_2| = 2|a_2 - b_2| \\ &\leq |a_2 - b_2| + |a_1 - b_1| + |a_1 + a_2 - b_1 - b_2| = \alpha. \end{aligned}$$

2. $a_1, b_1 \leq c_1$ and $c_2 \leq a_2, b_2$. Then it holds that:

$$\beta = |c_1 - c_1| + |c_2 - c_2| + |c_1 + c_2 - c_1 - c_2| = 0 \leq \alpha.$$

3. $a_1, b_1 \leq c_1$ and $b_2 \leq c_2 \leq a_2$:

$$\begin{aligned} \beta &= |c_1 - c_1| + |c_2 - b_2| + |c_1 + c_2 - c_1 - b_2| = 2|c_2 - b_2| \\ &\leq 2|a_2 - b_2| \leq \alpha. \end{aligned}$$

4. $c_1 \leq a_1, b_1$ and $c_2 \leq a_2, b_2$:

$$\begin{aligned} \beta &= |a_1 - b_1| + |c_2 - c_2| + |a_1 + c_2 - b_1 - c_2| = 2|a_1 - b_1| \\ &\leq |a_1 - b_1| + |a_2 - b_2| + |a_1 + a_2 - b_1 - b_2| = \alpha. \end{aligned}$$

5. $c_1 \leq a_1, b_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned}
\beta &= |a_1 - b_1| + |c_2 - b_2| + |a_1 + c_2 - b_1 - b_2| \\
&= \begin{cases} |b_1 - a_1| + (c_2 - b_2) + (a_1 - b_1) - (b_2 - c_2) & \text{if } a_1 - b_1 \geq b_2 - c_2 \\ |b_1 - a_1| + (c_2 - b_2) + (b_2 - c_2) - (a_1 - b_1) & \text{if } a_1 - b_1 < b_2 - c_2 \end{cases} \\
&= \begin{cases} |b_1 - a_1| + (a_1 - b_1) + 2(c_2 - b_2) & \text{if } a_1 - b_1 \geq b_2 - c_2 \\ 2|b_1 - a_1| & \text{if } a_1 - b_1 < b_2 - c_2 \end{cases} \\
&\leq \begin{cases} |b_1 - a_1| + (a_1 - b_1) + (c_2 - b_2) + (a_2 - b_2) & \text{if } a_1 - b_1 \geq b_2 - c_2 \\ |a_1 - b_1| + |a_2 - b_2| + |a_1 + a_2 - b_1 - b_2| & \text{if } a_1 - b_1 < b_2 - c_2 \end{cases} \\
&\leq \begin{cases} |b_1 - a_1| + (a_1 - b_1) + (a_2 - b_2) + (a_2 - b_2) & \text{if } a_1 - b_1 \geq b_2 - c_2 \\ |a_1 - b_1| + |a_2 - b_2| + |a_1 + a_2 - b_1 - b_2| & \text{if } a_1 - b_1 < b_2 - c_2 \end{cases} \\
&\leq \alpha.
\end{aligned}$$

6. $b_1 \leq c_1 \leq a_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned}
\beta &= |a_1 - c_1| + |c_2 - b_2| + |a_1 + c_2 - c_1 - b_2| \\
&= (a_1 - c_1) + (c_2 - b_2) + (a_1 - c_1) + (c_2 - b_2) \\
&= 2(a_1 - c_1) + 2(c_2 - b_2) \leq 2(a_1 - b_1) + 2(a_2 - b_2) \leq \alpha.
\end{aligned}$$

7. $b_1 \leq c_1 \leq a_1$ and $a_2 \leq c_2 \leq b_2$.

$$\begin{aligned}
\beta &= |a_1 - c_1| + |a_2 - c_2| + |a_1 + a_2 - c_1 - c_2| \\
&= \begin{cases} (a_1 - c_1) + (c_2 - a_2) + (a_1 - c_1) + (a_2 - c_2) & \text{if } a_1 - c_1 \geq c_2 - a_2 \\ (a_1 - c_1) + (c_2 - a_2) - (a_1 - c_1) - (a_2 - c_2) & \text{if } a_1 - c_1 < c_2 - a_2 \end{cases} \\
&= \begin{cases} 2(a_1 - c_1) \leq 2(a_1 - b_1) & \text{if } a_1 - c_1 \geq c_2 - a_2 \\ 2(c_2 - a_2) \leq 2(b_2 - a_2) & \text{if } a_1 - c_1 < c_2 - a_2 \end{cases} \\
&\leq \begin{cases} 2(a_1 - b_1) & \text{if } a_1 - c_1 \geq c_2 - a_2 \\ 2(b_2 - a_2) & \text{if } a_1 - c_1 < c_2 - a_2 \end{cases} \leq \alpha.
\end{aligned}$$

In the remaining cases, it is enough to exchange the roles of (a_1, b_1) , (a_2, b_2) and to apply the previous cases. \blacksquare

Lemma A.3 *If (a_1, a_2) , (b_1, b_2) and (c_1, c_2) are elements on $\mathcal{T} = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}$, then it holds that:*

$$\begin{aligned}
&|a_1 - b_1 - a_2 + b_2| + |a_1 - b_1| + |a_2 - b_2| \geq \\
&\quad |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| + \\
&\quad |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}|.
\end{aligned}$$

Proof: Let us consider some cases.

1. $a_1, b_1 \leq c_1$ and $a_2, b_2 \leq c_2$.

$$\begin{aligned} & |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\ & + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\ & = |c_1 - c_1 - a_2 + b_2| + |c_1 - c_1| + |a_2 - b_2| = 2|b_2 - a_2| \\ & \leq |a_1 - b_1 - a_2 + b_2| + |a_1 - b_1| + |a_2 - b_2|. \end{aligned}$$

2. $a_1, b_1 \leq c_1$ and $c_2 \leq a_2, b_2$.

$$\begin{aligned} & |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\ & + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\ & = |c_1 - c_1 - c_2 + c_2| + |c_1 - c_1| + |c_2 - c_2| = 0 \\ & \leq |a_1 - b_1 - a_2 + b_2| + |a_1 - b_1| + |a_2 - b_2|. \end{aligned}$$

3. $a_1, b_1 \leq c_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned} & |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\ & + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\ & = |c_1 - c_1 - c_2 + b_2| + |c_1 - c_1| + |c_2 - b_2| = 2|c_2 - b_2| \\ & \leq |a_1 - b_1 - a_2 + b_2| + |a_1 - b_1| + |a_2 - b_2|. \end{aligned}$$

4. $a_1, b_1 \leq c_1$ and $a_2 \leq c_2 \leq b_2$. It suffices to exchange the roles of (a_1, a_2) and (b_1, b_2) and to apply the previous case.

5. $c_1 \leq a_1, b_1$ and $a_2, b_2 \leq c_2$. Take (a_2, a_1) and (b_2, b_1) and apply case 2.

6. $c_1 \leq a_1, b_1$ and $c_2 \leq a_2, b_2$.

$$\begin{aligned} & |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\ & + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\ & = |a_1 - b_1 - c_2 + c_2| + |a_1 - b_1| + |c_2 - c_2| = 2|a_1 - b_1| \\ & = |a_1 - b_1 - a_2 + b_2| + |a_1 - b_1| + |a_2 - b_2|. \end{aligned}$$

7. $c_1 \leq a_1, b_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned} & |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\ & + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\ & = |a_1 - b_1 - c_2 + b_2| + |a_1 - b_1| + |c_2 - b_2| \\ & = \begin{cases} (a_1 - b_1) - (c_2 - b_2) + |a_1 - b_1| + (c_2 - b_2) & \text{if } a_1 - b_1 \geq c_2 - b_2 \\ 2(c_2 - b_2) - (a_1 - b_1) + |a_1 - b_1| & \text{if } a_1 - b_1 \leq c_2 - b_2 \end{cases} \\ & \leq \begin{cases} 2|a_1 - b_1| & \text{if } a_1 - b_1 \geq c_2 - b_2 \\ (a_2 - b_2) - (a_1 - b_1) + |a_1 - b_1| + |a_2 - b_2| & \text{if } a_1 - b_1 \leq c_2 - b_2 \end{cases} \\ & = |a_1 - b_1| + |a_2 - b_2| + |a_1 - b_1 - a_2 + b_2|. \end{aligned}$$

8. $c_1 \leq a_1, b_1$ and $a_2 \leq c_2 \leq b_2$. It suffices to exchange (a_1, a_2) and (b_1, b_2) and to apply the previous case.
9. $b_1 \leq c_1 \leq a_1$ and $a_2, b_2 \leq c_2$. It is enough to consider (a_2, a_1) and (b_1, b_2) and to apply case 3.
10. $b_1 \leq c_1 \leq a_1$ and $c_2 \leq a_2, b_2$. It suffices to consider (a_2, a_1) and (b_1, b_2) and to apply case 7.
11. $b_1 \leq c_1 \leq a_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |a_1 - c_1 - c_2 + b_2| + |a_1 - c_1| + |c_2 - b_2| \\
& = \begin{cases} 2(a_1 - c_1) + (c_2 - b_2) - (c_2 - b_2) & \text{if } a_1 - c_1 \geq c_2 - b_2 \\ (a_1 - c_1) + 2(c_2 - b_2) - (a_1 - c_1) & \text{if } a_1 - c_1 \leq c_2 - b_2 \end{cases} \\
& \leq \begin{cases} 2(a_1 - b_1) & \text{if } a_1 - c_1 \geq c_2 - b_2 \\ 2(a_2 - b_2) & \text{if } a_1 - c_1 \leq c_2 - b_2 \end{cases} \\
& = |a_1 - b_1| + |a_2 - b_2| + |a_1 - b_1 - a_2 + b_2|.
\end{aligned}$$

12. $b_1 \leq c_1 \leq a_1$ and $a_2 \leq c_2 \leq b_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |a_1 - c_1 - a_2 + c_2| + |a_1 - c_1| + |a_2 - c_2| \\
& = 2(a_1 - c_1) + 2(c_2 - a_2) \\
& \leq 2(a_1 - b_1) + 2(b_2 - a_2) \\
& = |a_1 - b_1| + |a_2 - b_2| + |a_1 - b_1 - a_2 + b_2|.
\end{aligned}$$

13. $a_1 \leq c_1 \leq b_1$. It is enough to consider (a_2, a_1) and (b_2, b_1) and to apply the previous cases. ■

Lemma A.4 *If (a_1, a_2) , (b_1, b_2) and (c_1, c_2) are elements on $\mathcal{T} = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}$, then:*

$$\begin{aligned}
& |a_1 - b_1| + |a_2 - b_2| + |a_1 - b_1 - a_2 + b_2| + |a_1 + a_2 - b_1 - b_2| \geq \\
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| + \\
& |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| + \\
& |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}|.
\end{aligned}$$

Proof: Throughout this proof we will use the fact that $|x+y|+|x-y| = \max\{2|x|, 2|y|\}$. Let us consider the following possibilities.

1. $a_1, b_1 \leq c_1$ and $a_2, b_2 \leq c_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |c_1 - c_1| + |a_2 - b_2| + |c_1 - c_1 - a_2 + b_2| + |c_1 - c_1 + a_2 - b_2| \\
& = 3|a_2 - b_2| \leq |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1| \\
& \leq |a_1 - b_1| + |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1|.
\end{aligned}$$

2. $a_1, b_1 \leq c_1$ and $c_2 \leq a_2, b_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |c_1 - c_1| + |c_2 - c_2| + |c_1 - c_1 - c_2 + c_2| + |c_1 - c_1 + c_2 - c_2| \\
& = 0 \leq |a_1 - b_1| + |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1|.
\end{aligned}$$

3. $a_1, b_1 \leq c_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |c_1 - c_1| + |c_2 - b_2| + |c_1 - c_1 - c_2 + b_2| + |c_1 - c_1 + c_2 - b_2| \\
& = 3|c_2 - b_2| \leq 3|a_2 - b_2| \\
& = |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1| \\
& \leq |a_1 - b_1| + |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1|.
\end{aligned}$$

4. $a_1, b_1 \leq c_1$ and $a_2 \leq c_2 \leq a_2$. It suffices to exchange the roles of (a_1, a_2) and (b_1, b_2) .

5. $c_1 \leq a_1, b_1$ and $a_2, b_2 \leq c_2$. It suffices to consider (a_2, a_1) and (b_2, b_1) and to apply case 2.

6. $c_1 \leq a_1, b_1$ and $c_2 \leq a_2, b_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |a_1 - b_1| + |c_2 - c_2| + |a_1 - b_1 + c_2 - c_2| + |a_1 - b_1 - c_2 + c_2| \\
& = 3|a_1 - b_1| \leq |a_1 - b_1| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1| \\
& \leq |a_1 - b_1| + |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1|.
\end{aligned}$$

7. $c_1 \leq a_1, b_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |a_1 - b_1| + |c_2 - b_2| + |a_1 - b_1 - c_2 + b_2| + |a_1 - b_1 + c_2 - b_2| \\
& = |a_1 - b_1| + |c_2 - b_2| + 2 \max(|a_1 - b_1|, |c_2 - b_2|) \\
& \leq |a_1 - b_1| + |a_2 - b_2| + 2 \max(|a_1 - b_1|, |a_2 - b_2|) \\
& \leq |a_1 - b_1| + |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1|.
\end{aligned}$$

8. $c_1 \leq a_1, b_1$ and $a_2 \leq c_2 \leq a_2$. It suffices to exchange the roles of (a_1, a_2) and (b_1, b_2) and to apply the previous case.

9. $b_1 \leq c_1 \leq a_1$ and $a_2, b_2 \leq c_2$. It is enough to consider (a_2, a_1) and (b_2, b_1) and to apply case 3.

10. $b_1 \leq c_1 \leq a_1$ and $c_2 \leq a_2, b_2$. Consider (a_2, a_1) and (b_2, b_1) and to apply case 7.

11. $b_1 \leq c_1 \leq a_1$ and $b_2 \leq c_2 \leq a_2$.

$$\begin{aligned}
& |\max\{a_1, c_1\} - \max\{b_1, c_1\}| + |\min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} - \min\{a_2, c_2\} + \min\{b_2, c_2\}| \\
& + |\max\{a_1, c_1\} - \max\{b_1, c_1\} + \min\{a_2, c_2\} - \min\{b_2, c_2\}| \\
& = |a_1 - c_1| + |c_2 - b_2| + |a_1 - c_1 - c_2 + b_2| + |a_1 - c_1 + c_2 - b_2| \\
& = |a_1 - c_1| + |c_2 - b_2| + 2 \max(|a_1 - c_1|, |c_2 - b_2|) \\
& \leq |a_1 - b_1| + |a_2 - b_2| + 2 \max(|a_1 - b_1|, |a_2 - b_2|) \\
& = |a_1 - b_1| + |a_2 - b_2| + |a_2 - b_2 - a_1 + b_1| + |a_2 - b_2 + a_1 - b_1|.
\end{aligned}$$

12. $b_1 \leq c_1 \leq a_1$ and $a_2 \leq c_2 \leq b_2$. It suffices to exchange the roles of (a_1, a_2) and (b_1, b_2) and to apply the previous case.

13. $a_1 \leq c_1 \leq b_1$. It suffices to exchange (a_1, a_2) and (b_1, b_2) and to apply the previous cases. ■

Lemma A.5 *If (a_1, a_2) , (b_1, b_2) and (c_1, c_2) are three elements in $\mathcal{T} = \{(x, y) \in [0, 1]^2 \mid x + y \leq 1\}$, then:*

$$\begin{aligned}
& |\max\{a_1 - 0.5, 0\} - \max\{b_1 - 0.5, 0\}| + \\
& |\max\{a_2 - 0.5, 0\} - \max\{b_2 - 0.5, 0\}| \geq \\
& |\max\{\max\{a_1, c_1\} - 0.5, 0\} - \max\{\max\{b_1, c_1\} - 0.5, 0\}| + \\
& |\max\{\min\{a_2, c_2\} - 0.5, 0\} - \max\{\min\{b_2, c_2\} - 0.5, 0\}|.
\end{aligned}$$

Proof: In order to prove this result, we are going to prove the following inequalities:

$$\begin{aligned}
& |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}| \geq \\
& |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}|, \\
& |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}| \geq \\
& |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}|,
\end{aligned}$$

for every $a, b, c \in [0, 1]$. Let us consider several cases.

1. $a \leq b \leq c$.

$$\begin{aligned} & |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}| \\ &= |\max\{c - 0.5, 0\} - \max\{c - 0.5, 0\}| \\ &= 0 \leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \\ & |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}| \\ &= |\max\{a - 0.5, 0\} + \max\{b - 0.5, 0\}|. \end{aligned}$$

2. $a \leq c \leq b$. This implies that $b - 0.5 \geq c - 0.5 \geq a - 0.5$, and therefore $\max\{b - 0.5, 0\} \geq \max\{c - 0.5, 0\} \geq \max\{a - 0.5, 0\}$.

$$\begin{aligned} & |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}| \\ &= |\max\{c - 0.5, 0\} - \max\{b - 0.5, 0\}| \\ &\leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}| \\ & |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}| \\ &= |\max\{a - 0.5, 0\} - \max\{c - 0.5, 0\}| \\ &\leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \end{aligned}$$

3. $b \leq a \leq c$.

$$\begin{aligned} & |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}| \\ &= |\max\{c - 0.5, 0\} - \max\{c - 0.5, 0\}| \\ &= 0 \leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \\ & |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}| \\ &= |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \end{aligned}$$

4. $b \leq c \leq a$. Then $a - 0.5 \geq c - 0.5 \geq b - 0.5$, and consequently $\max\{a - 0.5, 0\} \geq \max\{c - 0.5, 0\} \geq \max\{b - 0.5, 0\}$.

$$\begin{aligned} & |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}| \\ &= |\max\{a - 0.5, 0\} - \max\{c - 0.5, 0\}| \\ &\leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \\ & |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}| \\ &= |\max\{c - 0.5, 0\} - \max\{b - 0.5, 0\}| \\ &\leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \end{aligned}$$

5. $c \leq a \leq b$.

$$\begin{aligned} & |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}| \\ &= |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \\ & |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}| \\ &= |\max\{c - 0.5, 0\} - \max\{c - 0.5, 0\}| \\ &= 0 \leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \end{aligned}$$

6. $c \leq b \leq a$.

$$\begin{aligned}
 & |\max\{\max\{a, c\} - 0.5, 0\} - \max\{\max\{b, c\} - 0.5, 0\}| \\
 &= |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|. \\
 & |\max\{\min\{a, c\} - 0.5, 0\} - \max\{\min\{b, c\} - 0.5, 0\}| \\
 &= |\max\{c - 0.5, 0\} - \max\{c - 0.5, 0\}| \\
 &= 0 \leq |\max\{a - 0.5, 0\} - \max\{b - 0.5, 0\}|.
 \end{aligned}$$

Thus, for every $(a_1, a_2), (b_1, b_2), (c_1, c_2) \in \mathcal{T}$ it holds that:

$$\begin{aligned}
 & |\max\{a_1 - 0.5, 0\} - \max\{b_1 - 0.5, 0\}| + \\
 & |\max\{a_2 - 0.5, 0\} - \max\{b_2 - 0.5, 0\}| \geq \\
 & \quad |\max\{\max\{a_1, c_1\} - 0.5, 0\} - \max\{\max\{b_1, c_1\} - 0.5, 0\}| + \\
 & \quad |\max\{a_2 - 0.5, 0\} - \max\{b_2 - 0.5, 0\}| \geq \\
 & \quad |\max\{\max\{a_1, c_1\} - 0.5, 0\} - \max\{\max\{b_1, c_1\} - 0.5, 0\}| + \\
 & \quad |\max\{\min\{a_2, c_2\} - 0.5, 0\} - \max\{\min\{b_2, c_2\} - 0.5, 0\}|. \blacksquare
 \end{aligned}$$

List of symbols

(Ω, \mathcal{A}, P)	Probability space.
X, Y, Z, \dots	Random variables.
F_X, F_Y, F_Z, \dots	Cumulative distribution functions.
(Ω', \mathcal{A}')	Ordered space.
\succsim	Preference relation.
\succ	Strict preference relation.
\equiv	Indifference relation.
$\not\sim$	Incomparability relation.
\mathcal{D}	Set of random variables.
\succsim_{FSD}	First degree stochastic dominance.
\succsim_{SSD}	Second degree stochastic dominance.
\succsim_{nSD}	n -th degree stochastic dominance.
\mathcal{U}	Set of increasing functions $u : \mathbb{R} \rightarrow \mathbb{R}$.
\mathcal{U}^*	Set of increasing and bounded functions $u : \mathbb{R} \rightarrow \mathbb{R}$.
Q	Probabilistic relation.
\succsim_{SP}	Statistical preference.
$Q_{\frac{1}{2}}$	$\frac{1}{2}$ -cut of the probabilistic relation Q .
C	Copula.
M	Minimum copula.
W	Łukasiewicz copula.
π	Product copula.
φ	Generator of an Archimedean copula.
$\mathcal{L}(\Omega)$	Set of gambles on Ω .
\mathcal{K}	Set of gambles $\mathcal{K} \subseteq \mathcal{L}(\Omega)$.
\underline{P}	Lower prevision.

\overline{P}	Upper prevision.
$\mathcal{M}(\underline{P})$	Credal set associated with \underline{P} .
\underline{E}	Natural extension.
\underline{F}	Lower distribution function.
\overline{F}	Upper distribution function.
$(\underline{E}, \overline{F})$	P-box.
$\underline{P}_{(\underline{E}, \overline{F})}$	Lower probability associated with the p-box $(\underline{E}, \overline{F})$.
Π	Possibility measure.
π	Possibility distribution.
N	Necessity measure.
$[\delta, \pi]$	Cloud.
Γ	Random set.
$\Gamma^*(A)$	Upper inverse of Γ in A .
$S(\Gamma)$	Set of measurable selections of the random set Γ .
P_Γ^*	Upper probability defined from the random set Γ .
$P_{*\Gamma}$	Lower probability defined from the random set Γ .
$\mathcal{P}(\Gamma)$	Set of probabilities defined by the measurable selections.
$\beta_{[0,1]}$	Borel σ -algebra on $[0, 1]$.
$\lambda_{[0,1]}$	Lebesgue measure on $[0, 1]$.
$(C) \int f d\mu$	Choquet integral of f with respect to μ .
μ_A, ν_A	Membership and non-membership functions of an IF-set A .
π_A	Hesitation index of the IF-set A .
$[l_A, u_A]$	Lower and upper bounds of the IVF-set A .
$IFSS(\Omega)$	Set of all IF-sets defined on Ω .
$FS(\Omega)$	Set of all fuzzy sets defined on Ω .
$\mathcal{B}(p)$	Bernoulli distribution with parameter p .
$Exp(\lambda)$	Exponential distribution with parameter λ .
$\mathcal{U}(a, b)$	Uniform distribution in the interval (a, b) .
$\mathcal{P}_a(\lambda)$	Pareto distribution with parameter λ .
$\beta(p, q)$	Beta distribution with parameters p and q .
$\beta(p, q, a, b)$	Beta distribution on the interval (a, b) with parameters p and q .
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and variance σ^2 .
$\mathcal{N}(\vec{\mu}, \Sigma)$	Multidimensional normal distribution with vector of means $\vec{\mu}$ and matrix of variances-covariances Σ .

ρ	Correlation coefficient.
$Q_n(\cdot, [\cdot])$	General probabilistic relation.
δ_a	Dirac functional on the point a .
\succ_{pp}	Probabilistic prior relation.
\succ_{sd}	Strong dominance relation.
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \dots$	Sets of random variables.
\succeq_{FSD_i}	Imprecise first degree stochastic dominance.
$\mathcal{F}_X, \mathcal{F}_Y$	Sets of cumulative distribution functions.
bel	Belief function.
pl	Plausibility function.
\succeq_{SP_i}	Imprecise statistical preference.
$\mathbb{Q}_{\mathcal{X}, \mathcal{Y}}$	Profile of preferences of the sets of random variables \mathcal{X} and \mathcal{Y} .
Ω_1, Ω_2	Ordered spaces.
\underline{C}	Upper copula.
\overline{C}	Lower copula.
$(\underline{C}, \overline{C})$	Imprecise copula.
$\underline{P}_X \boxtimes \underline{P}_Y$	Independent natural extension of \underline{P}_X and \underline{P}_Y .
$\underline{P}_X \otimes \underline{P}_Y$	Strong product of \underline{P}_X and \underline{P}_Y .
\succeq_{uo}	Upper orthant relation.
\succeq_{lo}	Lower orthant relation.
D_{IFS}	IF-divergence.
l_{IFS}	Hamming distance for IF-sets.
d_{H}	Hausdorff distance for IF-sets.
D_C, D_L	Hong and Kim dissimilarities.
D_O	Li et al. dissimilarity.
D_{HB}	Mitchell dissimilarity.
D_e^p, D_h^p	Liang and Shi dissimilarities.
$D_{\text{HY}}^1, D_{\text{HY}}^2, D_{\text{HY}}^3$	Hung and Yang dissimilarities.
q_{IFS}	Euclidean distance for IF-sets.
D_s^p	Liang and Shi dissimilarity.
D_C	Chen dissimilarity.
D_{DC}	Dengfenf and Chuntian dissimilarity.
$D_{\omega 1}, D_{\text{pk}1}, D_{\text{pk}3}$	Hung and Yang dissimilarities.
h_{IFS}	Function that defines a local IF-divergence.
$CU([0, 1])$	Set of clouds defined on $[0, 1]$.
$IF^*([0, 1])$	Set of IF-sets on $[0, 1]$ such that $\mu_A(x) = 0$ and $\nu_A(y) = 0$ for some $x, y \in [0, 1]$.

Index

- Aggregation function, 199
- Aumann integral, 33, 204
- Belief function, 31, 166
- Binary relation, 9, 149
 - Incomparable relation, 9, 149
 - Indifference relation, 9, 149
 - Strict preference relation, 9, 149
 - Weak preference relation, 9, 149
- Bivariate distribution function, 210
- Choquet integral, 33, 204
- Cloud, 30, 348
- Coherent lower probability
 - n monotone, 225
- Comonotone functions, 32
 - Strict comonotone functions, 32, 333
- Comonotonic random variables, 23, 56, 57, 81, 83, 116, 134
- Concave order, 13, *see also* Second degree stochastic dominance
- Convergence, 15, 49
 - Convergence almost sure, 15, 49
 - Convergence in m -th mean, 15, 49
 - Convergence in distribution, 15, 49
 - Convergence in probability, 15, 49
- Copula, 21
 - Lukasiewicz copula, 21
 - Archimedean copula, 22
 - Imprecise copula, 228
 - Minimum copula, 21
 - Nilpotent Archimedean copula, 22, 70, 89
 - Product copula, 21
 - Strict Archimedean copula, 22, 67, 86, 123
- Countermonotonic random variables, 23, 57, 60, 82, 83
- Credal set, 24, 183, 231, 332, 349
- Dirac-delta functional, 115
- Distance for IF-sets, 262
- Envelope Theorem, 24, 225
- Finitely additive distribution functions, 178
- Fitness values, 130
- Fréchet-Hoeffding bounds, 22
- Fuzzy Sets, 34
- Fuzzy sets, 330
- Gamble, 24
- Gini Index, 246
- Hesitation index, 35, 332
- Hurwicz criterion, 164
- Idempotent aggregation function, 199
- IF-difference, 39, 262, 270–273
- IF-dissimilarity, 258
- IF-dissimilitude, 260
- IF-divergence, 259
- Imprecise probabilities, 24
- Interval-valued fuzzy sets, 34, 330
- Intuitionistic fuzzy sets, 34
- Intuitionistic fuzzy sets, 258
- Lorenz curves, 244, 359
- Lower distribution function, 26, 211, 334

- Lower prevision, 24
 - n monotone, 29
 - Coherent lower prevision, 24
- Lower probability, 25, 203, 209, 331
 - n monotone, 29
- Measurable selection, 31, 202, 332
- Median, 72, 194
 - Lower median, 209
 - Upper median, 209
- Monotone aggregation function, 199
- Multiattribute decision making, 250, 352

- Natural extension, 25, 233, 341
- Necessity measure, 29

- P-box, 26, 160, 334, 358
 - Bivariate p-box, 212
- Pareto dominance, 153
- Plausibility function, 31, 166
- Possibility measure, 29, 183, 330
- Preference structure, 10
 - Preference structure without incomparable elements, 10
- Probabilistic prior dominance, 131
- Probabilistic relation, 16, 200

- Random set, 31, 203, 331
- Rectangle inequality, 210
 - Imprecise rectangle inequalities, 217

- Self-Dual aggregation function, 199
- Sklar Theorem, 23
 - Imprecise Sklar Theorem, 232
- Statistical preference, 16
 - General statistical preference, 106
 - Imprecise statistical preference, 193
- Stochastic dominance, 10, 359
 - n -th degree stochastic dominance, 13
 - Bivariate stochastic dominance, 239
 - First degree stochastic dominance, 11
 - Imprecise stochastic dominance, 157
 - Second degree stochastic dominance, 12
- Stochastic order, 11, *see also* First degree stochastic dominance

- Strong dominance, 131
- Symmetric aggregation function, 199

- Upper distribution function, 26, 211, 334
- Upper prevision, 25
 - Coherent upper prevision, 25
- Upper probability, 25, 203, 209, 331

List of Figures

2.1	Example of first degree stochastic dominance: $X \succeq_{\text{FSD}} Y$	12
2.2	Example of second degree stochastic dominance: $X \succeq_{\text{SSD}} Y$	13
2.3	Interpretation of the reciprocal relation Q	16
2.4	Geometrical interpretation of the statistical preference: $X \succeq_{\text{SP}} Y$	17
2.5	Examples of the membership and non-membership functions of the IF-sets that express the P is a good politician (*), P is honest (◦) and P is close to the people (◊).	38
3.1	Graphics of the functions G_X^2 and G_Y^2	91
3.2	Graphics of the functions G_X^2 and G_Y^2	92
3.3	General relationship between stochastic dominance and statistical preference.	93
3.4	Probabilistic relation for the three dices.	104
3.5	Relationships among first and n^{th} degree stochastic dominance, statistical preference and the general statistical preference.	129
3.6	Summary of the relationships between strong dominance, first degree stochastic dominance, probabilistic prior and statistical preference given in Proposition 3.110.	133
3.7	Values of the probabilistic relation for different values of p . The above picture corresponds to intervals $[a_1, b_1] = [1, 4]$ and $[a_2, b_2] = [2, 3]$, and the picture below corresponds to intervals $[a_1, b_1] = [0.7, 1.4]$ and $[a_2, b_2] = [0.8, 1.2]$	142

- 4.1 Relationships among the different extensions of the binary relation for the comparison of sets of random variables. 151
- 4.2 Examples of several definitions of imprecise stochastic dominance. 158

- 5.1 Relationships among IF-divergences, IF-dissimilitudes, IF-dissimilarities and distances for IF-sets. 267

List of Tables

3.1	Definition of random variables X and Y	72
3.2	Definition of random variables X and Y	74
3.3	Characteristics of the continuous distributions to be studied.	95
3.4	Characterizations of statistical preference between independent random variables included in the same family of distributions.	102
3.5	Degrees of preference for the different values of the parameter p for $FMSE_1 = [1, 4]$ and $FMSE_2 = [2, 3]$	141
3.6	Degrees of preference for the different values of the parameter p for $FMSE_1 = [0.7, 1.4]$ and $FMSE_2 = [0.8, 1.2]$	141
3.7	Linguistic payoff matrix-Expert 1.	143
3.8	Linguistic payoff matrix-Expert 2.	143
3.9	Linguistic payoff matrix-Expert 3.	143
4.1	Summary of the properties of the binary relation \succeq that hold onto their extensions $\succeq_1, \dots, \succeq_6$	156
4.2	Quintiles of the Lorenz Curves associated with different countries.	244
4.3	Cumulative distribution functions associated with the Lorenz Curves of the countries.	245
4.4	Result of the comparison of the regions by means of the imprecise stochastic dominance.	245

4.5	Estimation of relative survival rates by cancer site. The rates are derived from SEER 1973-98 database, all ethnic groups, both sexes (except (*), only for male, and (**) for female). [191].	248
4.6	Estimation of relative mortality rates by cancer site.	249
4.7	Result of the comparison of the different groups of cancer by means of the imprecise stochastic dominance.	250
5.1	Behaviour of well-known dissimilarities and IF-divergences.	291
5.2	Six kinds of materials are represented by IF-sets.	351

Bibliography

- [1] F. Aiche and D. Dubois. An extension of stochastic dominance to fuzzy random variables. In *Proceedings of IPMU 2010*. Springer-Verlag, 2010.
- [2] M. A. Arcones, P. H. Kvam, and F. J. Samaniego. Nonparametric estimation of a distribution subject to a stochastic preference constraint. *Journal of the American Statistical Association*, 97:170–182, 2002.
- [3] B. Arnold. *Majorization and the Lorenz Order: A Brief Introduction*. Springer, 1987.
- [4] K. Atanassov. Intuitionistic fuzzy sets. In *Proceedings of VII ITKR, Sofia*, 1983.
- [5] K. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20:87–96, 1986.
- [6] K. Atanassov. Interval valued intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 31:343–349, 1989.
- [7] K. Atanassov. *Intuitionistic Fuzzy Sets: Theory and Applications*. Physica-Verlag, Wyrzburg, 1999.
- [8] K. Atanassov. Intuitionistic fuzzy sets: Past, present and future. In *Proceedings of the "European Society of Fuzzy Logic and Technology Conference" (EUSFLAT) 2003*, 2003.
- [9] K. Atanassov and G. Gargov. Intuitionistic fuzzy logic. *CR Acad. Bulg. Soc.*, 43(3):9–12, 1990.
- [10] K. Atanassov and G. Georgeiv. Intuitionistic fuzzy prolog. *Fuzzy Sets and Systems*, 53:121–128, 1993.
- [11] A. Atkinson. On the measurement of poverty. *Econometrica*, 55:749–764, 1987.
- [12] J. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30:445–462, 1962.

- [13] J. Aumann. Integral of set valued functions. *Journal of Mathematical Analysis and Applications*, 12:1–12, 1965.
- [14] B. De Baets and H. De Meyer. Transitivity frameworks for reciprocal relations: cycle-transitivity versus α -transitivity. *Fuzzy Sets and Systems*, 152(2):249 – 270, 2005.
- [15] B. De Baets, H. De Meyer, and K. De Loof. On the cycle-transitivity of the mutual rank probability relation of a poset. *Fuzzy Sets and Systems*, 161(20):2695 – 2708, 2010.
- [16] B. De Baets, H. De Meyer, B. De Schuymer, and S. Jenei. Cyclic evaluation of transitivity of reciprocal relations. *Social Choice and Welfare*, 26:217–238, 2006.
- [17] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.
- [18] F. Belzunce, J. A. Mercader, J. M. Ruiz, and F. Spizzichino. Stochastic comparisons of multivariate mixture models. *Journal of Multivariate Analysis*, 100:1657–1669, 2009.
- [19] S. Benferhat, D. Dubois, and H. Prade. Towards a possibilistic logic handling of preferences. *Applied Intelligence*, 14:303–317, 2001.
- [20] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 1985.
- [21] J. C. Bezdek, D. Spliman, and R. Spilman. A fuzzy relation space for group decision theory. *Fuzzy Sets and Systems*, 1:255–268, 1978.
- [22] B. Bhandari, N. R. Pal, and D. D. Majumder. Fuzzy divergence, probability measure of fuzzy events and image thresholding. *Pattern Recognition Letters*, 13:857–867, 1992.
- [23] P. K. Bhatia and S. Singh. A new measure of fuzzy directed divergence and its application in image segmentation. *Journal of Intelligent Systems and Applications*, 4:81–89, 2013.
- [24] P. Billingsley. *Probability and Measure*. Wiley-Interscience, 3 edition, April 1995.
- [25] P. J. Boland, M. Hollander, K. Joag-Dev, and S. Kochar. Bivariate dependence properties of order statistics. *Journal of Multivariate Analysis*, 56:75–89, 1996.
- [26] P. J. Boland, H. Singh, and B. Cukic. The stochastic precedence ordering with applications in sampling and testing. *Journal of Applied Probability*, 41:73–82, 2004.
- [27] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems*, 84:143–153, 1996.

- [28] H. Brenner. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. *The Lancet*, 360:1131–1135, 2002.
- [29] P. Burrillo and H. Bustince. Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 78(3):305–316, 1996.
- [30] H. Bustince and P. Burrillo. Vague sets are intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 79:403–405, 1996.
- [31] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. *Aggregation operators, New Trends and Applications*. Physica-Verlag, Heidelberg, 2002.
- [32] A. Casero. Uso de simulación para el estudio de la preferencia estadística entre variables aleatorias. Trabajo Académicamente dirigido, Licenciatura de Matemáticas, Dirigido por E. Miranda, 2011.
- [33] D. Cayrac, D. Dubois, M. Haziza, and H. Prade. Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application. *IEEE Transactions on Fuzzy Systems*, 4:251–259, 1996.
- [34] T. Chaira and A. K. Ray. Segmentation using fuzzy divergence. *Pattern Recognition Letters*, 24:1837–1844, 2003.
- [35] A. Chateauneuf, M. Cohen, and R. Kast. Comonotone random variables in economics: a review of some results. *Cahiers d'ECOMATH*, 97.32, 1997.
- [36] S. M. Chen. Measures of similarity between vague sets. *Fuzzy Sets and Systems*, 67:217–223, 1995.
- [37] S. M. Chen. Similarity measures between vague sets and between elements. *IEEE Trans. Syst. Mn Cybernet.*, 27(1):153–158, 1997.
- [38] R. Cheng, B. Evans, and J. Williams. Confidence band estimation for distributions used in probabilistic design. *International Journal of Mechanical Sciences*, 30:835–845, 1988.
- [39] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.
- [40] M. Condorcet. *An essay on the application of probability theory to plurality decision making: an election between three candidates*. Reprinted in 1989, F. Sommerlad, I. McLean (Eds.), 1785.
- [41] O. Cerdón, F. Herrera, F. Gomide, F. Hoffmann, and L. Magdalena. Ten years of genetic-fuzzy systems: a current framework and new trends. In *Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference, Vancouver-Canada*, pages 1241–1246. 2001.

- [42] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific Publishing, 2001.
- [43] I. Couso and D. Dubois. An imprecise probability approach to joint extensions of stochastic and interval orderings. In *Proceedings of IPMU 2012*. Springer-Verlag, 2012.
- [44] I. Couso, L. Garrido, S. Montes, and L. Sánchez. Expected pairwise comparison of the values of a fuzzy random variable. In C. Borgelt, G. González-Rodríguez, W. Trutschnig, M.A. Lubiano, M.A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77 of *Advances in Intelligent and Soft Computing*, pages 105–114. Springer, 2010.
- [45] I. Couso, L. Garrido, and L. Sánchez. Similarity and dissimilarity measures between fuzzy sets: A formal relational study. *Information Sciences*, 229(0):122 – 141, 2013.
- [46] I. Couso and L. Sánchez. The behavioural meaning of the median. In C. Borgelt, G. González-Rodríguez, W. Trutschnig, M.A. Lubiano, M.A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77 of *Advances in Intelligent and Soft Computing*, pages 115–122. Springer, 2010.
- [47] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.
- [48] S. K. De, R. Biswas, and A. R. Roy. An application of intuitionistic fuzzy sets in medical diagnosis. *Fuzzy Sets and Systems*, 117:209–213, 2001.
- [49] B. De Baets and H. De Meyer. On the cycle-transitive comparison of artificiall coupled random variables. *International Journal of Approximate Reasoning*, 47:306–322, 2008.
- [50] G. de Cooman. Integration in possibility theory. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals – Theory and Applications*, pages 124–160. Physica-Verlag (Springer), Heidelberg, 2000.
- [51] G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.
- [52] G. de Cooman, E. Miranda, and M. Zaffalon. Independent natural extension. *Artificial*, 175(12–13):1911–1950, 2011.
- [53] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n -Monotone exact functionals. *Journal of Mathematical Analysis and Applications*, 347:143–156, 2008.

- [54] H. De Meyer, B. De Baets, and B. De Schuymer. On the transitivity of the comonotonic and countermonotonic comparison of random variables. *Journal of Multivariate Analysis*, 98:177–193, 2007.
- [55] B. De Schuymer, H. De Meyer, and B. De Baets. *A fuzzy approach to stochastic dominance of random variables*, pages 253–260. Lecture Notes in Artificial Intelligence 2715, 2003.
- [56] B. De Schuymer, H. De Meyer, and B. De Baets. Cycle-transitivity comparison of independent random variables. *Journal of Multivariate Analysis*, 96:352–373, 2005.
- [57] B. De Schuymer, H. De Meyer, B. De Baets, and S. Jenei. On the cycle-transitivity of the dice model. *Theory and Decision*, 54:261–285, 2003.
- [58] B. De Schuymer, H. De Meyer, and B. De Baets. Extreme copulas and the comparison of ordered lists. *Theory and Decision*, 62:195–217, 2007.
- [59] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [60] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic, Dordrecht, 1994.
- [61] T. Denoeux. Extending stochastic ordering to belief functions on the real line. *Information Sciences*, 179:1362–1376, 2009.
- [62] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial Theory for Dependent Risks Measures, Orders and Models*. John Wiley & Sons, 2005.
- [63] G. Deschrijver and E. Kerre. A generalization of operators on intuitionistic fuzzy sets using triangular norms and conorms. *Notes on Intuitionistic Fuzzy Sets*, 1:19–27, 2002.
- [64] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations- ii: Generalised p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.
- [65] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: Ii. clouds. *International Journal of Approximate Reasoning*, 49(3):664–677, 2008.
- [66] P. Dirac. *Principles of quantum mechanics*. Oxford at the Clarendon Press, 4th edition, 1958.
- [67] D. Dubois, H. Fargier, and P. Perny. Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach. *Artificial Intelligence*, 148:219–260, 2003.

- [68] D. Dubois, S. Gottwald, P. Hajek, J. Kacprzyk, and H. Prade. Terminological difficulties in fuzzy set theory—the case of “intuitionistic fuzzy sets”. *Fuzzy Sets and Systems*, 156:485–491, 2005.
- [69] D. Dubois, E. Kerre, and R. Mesiar. Fuzzy interval analysis. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy sets*, pages 483–581. Kluwer Academic Publishers, Boston, 2000.
- [70] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.
- [71] D. Dubois and H. Prade. *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.
- [72] D. Dubois and H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets and Systems*, 160(10):1355–1366, 2009.
- [73] D. Dubois and H. Prade. Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems*, 192:3–24, 2012.
- [74] J. Fan and W. Xie. Distance measure and induced fuzzy entropy. *Fuzzy Sets and Systems*, 104:305–314, 1999.
- [75] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003.
- [76] S. Ferson and W. Tucker. Probability boxes as info-gap models. In *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*, New York (USA), 2008.
- [77] M. Finkelstein. On stochastic comparisons of population densities and life expectancies. *Demographic Research*, 13:143–162, 2005.
- [78] P. C. Fishburn. *Interval Orders and Interval Graphs*. Wiley, New York, 1985.
- [79] M. Fréchet. Généralisations du théorème des probabilités totales. *Fundamenta Mathematicae*, 25:379–387, 1935.
- [80] J. L. García-Lapresta and B. Llamazares. Majority decisions based on differences of votes. *Journal of Mathematical Economics*, 35:463–381, 2001.
- [81] I. Georgescu. *Fuzzy Choice Functions, a Revealed Preference Approach*. Springer, Berlin, 2007.
- [82] I. Gilboa and D. Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

- [83] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. Valencia University Press, Valencia, 1980.
- [84] I. R. Goodman. Fuzzy sets as equivalence classes of random sets. In R. R. Yager, editor, *Fuzzy Sets and Possibility Theory*, pages 327–342. Pergamon Press, Oxford, 1982.
- [85] P. Grzegorzewski. Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the hausdorff metric. *Fuzzy Sets and Systems*, 148:319–328, 2004.
- [86] P. Grzegorzewski and E. Mrowka. Probability of intuitionistic fuzzy events. In P. Grzegorzewski, O. Hryniewicz, and M. A. Gil, editors, *Soft methods in probability, statistics and data analysis*, pages 105–115. Physica-Verlag, 2002.
- [87] F. Herrera. Genetic fuzzy systems: taxonomy and current research trend and prospects. *Evolutionary Intelligence*, 1:27–46, 2008.
- [88] C. J. Himmelberg. Measurable relations. *Fundamenta Mathematicae*, 87:53–72, 1975.
- [89] D. H. Hong and C. Kim. A note on similarity measures between vague sets and between elements. *Information Sciences*, 115:83–96, 1999.
- [90] T. Hu and C. Xie. Negative dependence in the balls and binds experiment with applications to order statistics. *Journal of Multivariate Analysis*, 97:1342–1354, 2006.
- [91] E. Hüllermeier. *Cased-based approximate reasoning*. Springer, 2007.
- [92] W-L. Hung and M-S. Yang. Similarity measures of intuitionistic fuzzy sets based on hausdorff distances. *Pattern Recognition Letters*, 25:1603–1611, 2004.
- [93] W-L. Hung and M-S. Yang. Similarity measures of intuitionistic fuzzy sets based on l_p metric. *International Journal of Approximate Reasoning*, 46:120–136, 2007.
- [94] W-L. Hung and M-S. Yang. On similarity measures between intuitionistic fuzzy sets. *Internation Journal of Intelligent Systems*, 23:364–383, 2008.
- [95] M. Islam and J. Braden. Bio-economics development of floodplains: Farming versus fishing in bangladesh. *Environment and Development Economics*, 11(1):95–126, 2006.
- [96] J.-Y. Jaffray and M. Jeleva. Information processing under imprecise risk with the Hurwicz criterion. In *Proceedings of the Fifth Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA*, 2007.

-
- [97] V. Janiš. T-norm based cuts of intuitionistic fuzzy sets. *Information Sciences*, 180(7):1134–1137, 2010.
- [98] A. Kaur, L. S. Prakasa Rao, and H. Singh. Testing for second-order stochastic dominance of two distributions. *Econometric Theory*, 10:849–866, 1994.
- [99] E. P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2000.
- [100] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, Dordrecht, 2000.
- [101] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, New Jersey, 1995.
- [102] V. Krätschmer. When fuzzy measures are upper envelopes of probability measures. *Fuzzy Sets and Systems*, 138:455–468, 2003.
- [103] E. Kriegler. Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning*, 39(2–3):185–209, 2005.
- [104] R. Kruse and K. D. Meyer. *Statistics with vague data*. D. Reidel Publishing Company, Dordrecht, 1987.
- [105] H. Kwakernaak. Fuzzy random variables, i, ii. *Information Sciences*, 15-17:253–178, 1979.
- [106] E. L. Lehmann. Ordering families of distributions. *Annals of Mathematical Statistics*, 26:399–419, 1955.
- [107] I. Levi. *The enterprise of knowledge*. MIT Press, Cambridge, 1980.
- [108] H. Levy. Stochastic dominance and expected utility: survey and analysis. *Management Science*, 38:555–593, 1992.
- [109] H. Levy. *Stochastic Dominance*. Kluwer Academic Publishers, 1998.
- [110] M. Levy and H. Levy. Testing for risk aversion: a stochastic dominance approach. *Economics Letters*, 71:233–240, 2001.
- [111] D. F. Li and C. T. Chen. New similarity measures of intuitionistic fuzzy sets and applications to pattern recognitions. *Pattern Recognition Letters*, 23:221–225, 2002.
- [112] X. Li and X. Hu. Some new stochastic comparisons for redundancy allocations in series and parallel systems. *Statistics and Probability Letters*, 78:3388–3394, 2008.
- [113] Y. Li, D. L. Olson, and Z. Qin. Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis. *Pattern Recognition Letters*, 28:278–285, 2007.

- [114] Z. Liang and P. Shi. Similarity measures on intuitionistic fuzzy sets. *Pattern Recognition Letters*, 24:2687–2693, 2003.
- [115] C.J. Liau. On the possibility theory-based semantics for logics of preference. *International Journal of Approximate Reasoning*, 20:173–190, 1999.
- [116] P. Limbourg. Multiobjective optimization of problems with epistemic uncertainty. In *Proceedings of Third International Conference on Evolutionary Multi-Criterion Optimization, Guanajuato-Mexico*, pages 413–427, 2005.
- [117] M. Lorenz. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(209–219):70, 1905.
- [118] P. Lui, F. Jin, X. Zhang, Y. Su, and M. Wang. Research on the multi-attribute decision-making under risk with interval probability based on prospects theory and the uncertain linguistic variables. *Knowledge-Based Systems*, 24:554–561, 2011.
- [119] X. Lui. Entropy, distance measure and similarity measure of fuzzy sets and their relations. *Fuzzy Sets and Systems*, 52:305–318, 1992.
- [120] D. Martinetti, V. Janiš, and S. Montes. Cuts of intuitionistic fuzzy sets respecting fuzzy connectives. *Information Sciences*, 232:267–275, 2013.
- [121] D. Martinetti, I. Montes, and S. Díaz. Transitivity of the fuzzy and the probabilistic relations associated to a set of random variables. In P. Burrillo, H. Bustince, B. De Baets, and J. Fodor, editors, *Proceedings of EUROFUSE 2009 Conference*, pages 21–26, 2009.
- [122] D. Martinetti, I. Montes, S. Díaz, and S. Montes. A study of the transitivity of the probabilistic and fuzzy relation. *Fuzzy Sets and Systems*, 184:156–170, 2011.
- [123] J. M. Merigó and A. M. Gil-Lafuente. Induced 2-tuple linguistic generalized aggregation operators and their application in decision-making. *Information Sciences*, 236:1–16, 2013.
- [124] P. Mikusiński, H. Sherwook, and M.D. Taylor. The Fréchet bounds revisited. *Real Anal Exchange*, 17:759–764, 1991-92.
- [125] E. Miranda. *Análisis de la información probabilística de los conjuntos aleatorios*. PhD thesis, University of Oviedo, 2003. In Spanish.
- [126] E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.
- [127] E. Miranda, I. Couso, and P. Gil. Upper probabilities and selectors of random sets. In P. Grzegorzewski, O. Hryniewicz, and M. A. Gil, editors, *Soft methods in probability, statistics and data analysis*, pages 126–133. Physica-Verlag, Heidelberg, 2002.

- [128] E. Miranda, I. Couso, and P. Gil. A random set characterization of possibility measures. *Information Sciences*, 168:51–75, 2004.
- [129] E. Miranda, I. Couso, and P. Gil. Random intervals as a model for imprecise information. *Fuzzy Sets and Systems*, 154:386–412, 2005.
- [130] E. Miranda, I. Couso, and P. Gil. Approximations of upper and lower probabilities by measurable selections. *Information Sciences*, 180:1407–1417, 2010.
- [131] E. Miranda, I. Couso, and P. Gil. Upper probabilities attainable by distributions of measurable selections. *Information Sciences*, 180(8):107–1417, 2010.
- [132] E. Miranda, G. de Cooman, and I. Couso. Lower previsions induced by multi-valued mappings. *Journal of Statistical Planning and Inference*, 133(1):173–197, 2005.
- [133] E. Miranda, G. de Cooman, and E. Quaeghebeur. Finitely additive extensions of distribution functions and moment sequences: The coherent lower prevision approach. *International Journal of Approximate Reasoning*, 48(1):132–155, 2008.
- [134] E. Miranda and I. Montes. Imprecise preferences by means of probability boxes. In *Proceedings of 4th ERCIM*, page 63, 2011.
- [135] E. Miranda, I. Montes, R. Pelessoni, and P. Vicig. Bivariate p-boxes. In *Proceedings of 8th ISIPTA*, 2013.
- [136] E. Miranda, M. Troffaes, and S. Destercke. Generalised p-boxes on totally ordered spaces. In D. Dubois, M. Lubiano, H. Prade, M. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods for Handling Variability and Imprecision*, pages 235–242. Springer, 2008.
- [137] E. Miranda and M. Zaffalon. Coherence graphs. *Artificial Intelligence*, 173(1):104–144, 2009.
- [138] H. B. Mitchell. On the dengfeng–chuntian similarity measure and its application to pattern recognition. *Pattern Recognition Letters*, 24:3101–3104, 2003.
- [139] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley, 2002.
- [140] I. Montes. Some general comments about statistical preference. In *Proceedings of I TUIM Conference*, 2010.
- [141] I. Montes. Divergences for intuitionistic fuzzy sets. In *Proceedings of II TUIM Conference*, 2011.
- [142] I. Montes. Comparison of more than two random variables by means of the statistical preference. In *Proceedings of XXXIII SEIO Conference*, 2012.

- [143] I. Montes, S. Díaz, and S. Montes. Comparison of imprecise fitness values modelled by beta distributions. pages 489–494, 2010. cited By (since 1996)0.
- [144] I. Montes, J. Hernández, D. Martinetti, and S. Montes. Characterization of continuous t-norms compatible with zadeh’s probability of fuzzy events. *Fuzzy Sets and Systems*, 228:29–43, 2013.
- [145] I. Montes, V. Janiš, and S. Montes. An axiomatic definition of divergence for intuitionistic fuzzy sets. volume 1, pages 547–553, 2011. cited By (since 1996)0.
- [146] I. Montes, V. Janiš, and S. Montes. On the study of some measures of comparison of if-sets. In *Proceedings of XVI ESTYLF Conference*, 2011.
- [147] I. Montes, V. Janiš, and S. Montes. Local if-divergences. *Communications in Computer and Information Science*, 298 CCIS(PART 2):491–500, 2012. cited By (since 1996)0.
- [148] I. Montes, D. Martinetti, S. Díaz, and S. Montes. A characterisation of statistical preference by means of the median. 2012. Submitted for publication.
- [149] I. Montes, D. Martinetti, and S. Díaz. On the statistical preference as a pairwise comparison for random variables. In *Proceedings of EUROFUSE 2009 Conference*, 2009.
- [150] I. Montes, D. Martinetti, S. Díaz, and S. Montes. Comparison of random variables coupled by archimedean copulas. In Christian Borgelt, Gil González-Rodríguez, Wolfgang Trutschnig, María Lubiano, María Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77 of *Advances in Intelligent and Soft Computing*, pages 467–474. Springer Berlin / Heidelberg, 2010.
- [151] I. Montes, D. Martinetti, S. Díaz, and S. Montes. Estudio de la preferencia estadística en distribuciones normales bidimensionales. In *Proceedings of XXXII SEIO Conference*, 2010.
- [152] I. Montes, D. Martinetti, S. Díaz, and S. Montes. Statistical preference as a comparison method of two imprecise fitness values. In *Proceedings of XV ESTYLF Conference*, 2010.
- [153] I. Montes, D. Martinetti, S. Díaz, and S. Montes. Statistical preference as a tool in consensus processes. In Enrique Herrera-Viedma, José García-Lapresta, Janusz Kacprzyk, Mario Fedrizzi, Hannu Nurmi, and Slawomir Zadrozny, editors, *Consensual Processes*, volume 267 of *Studies in Fuzziness and Soft Computing*, pages 65–92. Springer Berlin / Heidelberg, 2011.
- [154] I. Montes, D. Martinetti, and S. Montes. Statistical preference for testing algorithms. In *Proceedings of Centennial Congress of the Spanish Royal Mathematical Society RSME*, 2011.

- [155] I. Montes, E. Miranda, and S. Díaz. Preferences with p-boxes. In *Proceedings of 4th WPMIIP*, 2011.
- [156] I. Montes, E. Miranda, and S. Montes. Decision making with imprecise probabilities and utilities by means of statistical preference and stochastic dominance. *European Journal of Operational Research*, 234(1):209–220, 2014.
- [157] I. Montes, E. Miranda, and S. Montes. Stochastic dominance with imprecise information. *Computational Statistics and Data Analysis*, 71/C:867–885, 2014. In press.
- [158] I. Montes, N. R. Pal, V. Janiš, and S. Montes. Divergence measures for intuitionistic fuzzy sets. *IEEE Transactions on Fuzzy Systems*, Accepted to publication, 2014.
- [159] S. Montes. *Partitions and divergence measures in fuzzy models*. PhD thesis, University of Oviedo, 1998.
- [160] S. Montes, I. Couso, P. Gil, and C. Bertoluzza. Divergence measure between fuzzy sets. *International Journal of Approximate Reasoning*, 30:91–105, 2002.
- [161] S. Montes, T. Iglesias, V. Janiš, and I. Montes. A common framework for some comparison measures of if-sets. In *Proceedings of IWIFSGN Conference*, 2012.
- [162] S. Montes, D. Martinetti, I. Montes, and S. Díaz. Graded comparison of imprecise fitness values. In *Proceedings of Hard & Soft Conference*, 2010.
- [163] S. Montes, D. Martinetti, I. Montes, and S. Díaz. Min-transitivity of graded comparisons for random variables. In *Proceedings of FUZZ-IEEE 2010, Barcelona-Spain*. 2010.
- [164] R. F. Nau. Indeterminate probabilities on finite sets. *The Annals of Statistics*, 20:1737–1767, 1992.
- [165] R. F. Nau. The shape of incomplete preferences. *Annals of Statistics*, 34(5):2430–2448, 2006.
- [166] R. Nelsen. *An introduction to copulas*. Springer, New York, 1999.
- [167] R. Nelsen, J. J. Quesada Molina, J. A. Rodríguez-Lallena, and M. Úbeda Flores. Best-possible bounds on sets of bivariate distribution functions. *Journal of Multivariate Analysis*, 90:348–358, 2004.
- [168] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.
- [169] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65(3):531–542, 1978.
- [170] H. T. Nguyen. *An introduction to random sets*. Chapman and Hall, 2006.

- [171] N. Noyan, G. Rudolf, and A. Ruszczyński. Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research L*, 34:653–659, 2006.
- [172] M. Oberguggenberger and W. Fellin. Reliability bounds through random sets: non-parametric methods and geotechnical applications. *Computers and Structures*, 86(10):1093–1101, 2008.
- [173] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations and risk measures. *European Journal of Operational Research*, 116:33–50, 1999.
- [174] G. Owen. *Game theory*. Academic Press, San Diego. Third Edition, 1995.
- [175] N. R. Pal and S. K. Pal. Entropy, a new definition and its applications. *IEEE Trans. Syst. Mn Cybernet.*, 21(5):1260–1270, 1991.
- [176] R. Pelessoni, P. Vicig, I. Montes, and E. Miranda. Imprecise copulas and bivariate stochastic orders. In *Proceedings of EUROFUSE 2013 Conference*, 2013.
- [177] E. Raufaste, R. Da Silva Neves, and C. Mariné. Testing the descriptive validity of possibility theory in human judgements of uncertainty. *Artificial Intelligence*, 148:197–218, 2003.
- [178] D. Ríos Insua. On the foundations of decision analysis with partial information. *Theory and Decision*, 33:83–100, 1992.
- [179] M. Roubens and Ph. Vincke. *Preference Modelling*. Springer-Verlag, Berlin, 1985.
- [180] L. Sánchez and I. Couso. Advocating the use of imprecise observed data in genetic fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 15:551–562, 2007.
- [181] L. Sánchez, I. Couso, and J. Casillas. A multiobjective genetic fuzzy system with imprecise probability fitness for vague data. In *Proceedings of 2006 IEEE International Conference on Evolutionary Fuzzy Systems, Ambleside-UK*. 2006.
- [182] L. Sánchez, I. Couso, and J. Casillas. Modeling vague data with genetic fuzzy system under a combination of crisp and imprecise criteria. In *Proceedings of 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, Honolulu-USA*. 2007.
- [183] L. Sánchez, I. Couso, and J. Casillas. Genetic learning of fuzzy rules based on low quality data. *Fuzzy Sets and Systems*, 160:2524–2552, 2009.
- [184] J. K. Satia and R. E. Lave. Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21:728–740, 1973.

- [185] M. Scarsini. Copulae of capacities on product spaces. In Ludger Rüschendorf, Berthold Schweizer, and Michael Dee Taylor, editors, *Distributions with Fized Marginals and Related Topics*, volume 28 of *IMS Lecture Notes- Monograph Series*, pages 307–318. Institute of Mathematical Statistics, 1996.
- [186] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. A representation of partially ordered preferences. *The Annals of Statistics*, 23:2168–2217, 1995.
- [187] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
- [188] M. Shaked and J. G. Shanthikumar. *Stochastic Orders and their applications*. Springer, 2006.
- [189] A. Sklar. Fonctions de répartition à n-dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.
- [190] C. Strobl, A. L. Boulesteix, and T. Augustin. Unbiased split selection for classification trees based on the gini index. *Computational Statistics and Data Analysis*, 52(1):483–501, 2007.
- [191] Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-1998). *Bethesda (MD): National Cancer Institute, DCCPS, Cancer Surveillance Reasearch Program, Cancer Statistics Branch, released April 2001, based on the August 2000 submission*.
- [192] R. Szekli. *Stochastic ordering and dependence in applied probability*. Springer, 1995.
- [193] E. Szmidt and J. Kacprzyk. Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 114:505–518, 2000.
- [194] E. Szmidt and J. Kacprzyk. An application of intuitionistic fuzzy set similarity measures to a multi-criteria decision making problem. In Leszek Rutkowski, Ryszard Tadeusiewicz, LotfiA. Zadeh, and JacekM. Å»urada, editors, *Artificial Intelligence and Soft Computing ICAISC 2006*, volume 4029 of *Lecture Notes in Computer Science*, pages 314–323. Springer Berlin Heidelberg, 2006.
- [195] E. Szmidt and J. Kacprzyk. Analysis of similarity measures for atannasov’s intuitionistic fuzzy sets. In J. P. Carvalho, D. Dubois, U. Kaymak, and J. M. C. Sousa, editors, *Proceedings of “International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference” (IFSA-EUSFLAT 2009)*, pages 1416–1421. 2009.
- [196] H. Tanaka and P. J. Guo. *Possibilistic Data Analysis for Operations Research*. Physica-Verlag, Heidelberg, 1999.
- [197] J. Teich. Pareto-front exploration with uncertain objectives. In *Proceedings of First International Conference on Evolutionary Multi-Criterion Optimization*. 2001.

- [198] M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modell. *International Journal of Approximate Reasoning*, 52(6):767–791, 2011.
- [199] M. Troffaes, E. Miranda, and S. Destercke. On the connection between probability boxes and possibility measures. In *Proceedings of EUSFLAT'2011*, pages 836–843, Aix-les-Bains, France, 2011.
- [200] M. Troffaes, E. Miranda, and S. Destercke. On the connection between probability boxes and possibility measures. *Information Sciences*, 224:88–108, 2013.
- [201] M. C. M. Troffaes. *Optimality, Uncertainty, and Dynamic Programming with Lower Previsions*. PhD thesis, Ghent University, Ghent, Belgium, March 2005.
- [202] M. C. M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29, 2007.
- [203] P. Wakker. Characterizing optimism and pessimism through comonotonicity. *Journal of Economic Theory*, 52:453–463, 1990.
- [204] P. Walley. Coherent lower (and upper) probabilities. Technical Report Statistics Research Report 22, University of Warwick, Coventry, 1981.
- [205] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [206] W. Wang and X. Xin. Distance measure between intuitionistic fuzzy sets. *Pattern Recognition Letters*, 26:2063–2069, 2005.
- [207] X. Wang, C. Wu, and X. Wu. Choice functions in fuzzy environment: An overview. *Studies in Fuzziness and Soft Computing*, 261:149–169, 2010. cited By (since 1996)2.
- [208] G. A. Whitmore and M. C. Findlay. *Stochastic dominance: An approach to decision-making under risk*. Lexington Books, 1978.
- [209] R. Williamson. *Probabilistic arithmetic*. PhD thesis, University of Queensland, 1989.
- [210] C. Wrather and P. L. Yu. Probability dominance in random outcomes. *Journal of Optimization Theory and Applications*, 36:315–334, 1982.
- [211] Z. Xu. Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making. *Fuzzy Optimization and Decision Making*, 6:109–121, 2007.
- [212] Z. S. Xu and J. Chen. An overview of distance and similarity measures of intuitionistic fuzzy sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(4):529–555, 2008.

-
- [213] Y. Yuan and H. Zhuang. A general algorithm for generating fuzzy classification rules. *Fuzzy Sets and Systems*, 84:1–19, 1996.
- [214] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
- [215] L. A. Zadeh. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23:421–427, 1968.
- [216] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes interval-valued fuzzy sets. *IEEE Trans. Syst. Mn Cybernet.*, 3:28–44, 1973.
- [217] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.
- [218] M. Zaffalon and E. Miranda. Conservative inference rule for uncertain reasoning under incompleteness. *Journal of Artificial Intelligence Research*, 34:757–821, Jan 2009.
- [219] M. Zaffalon, K. Wesnes, and O. Petrini. Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artificial Intelligence in Medicine*, 29:61–79, 2003.