UNIVERSITEIT GENT

FACULTEIT WETENSCHAPPEN

Bell inequalities in cardinality-based similarity measurement

Saskia Janssens

Supervisors: prof. dr. Bernard De Baets
prof. dr. Hans De Meyer

Dissertation submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) in Sciences, Mathematics

Academic year 2005–2006

To my parents, who encouraged me to dream ...

and to Bart & Lennert, who made so many of my dreams come true.

Dankwoord

Het is me gelukt... Heel gelukkig ben ik en ook fier dat mijn boekje nu naast dat van Bart kan komen te staan in onze boekenkast. Een heleboel mensen - het is een onbegonnen taak om deze allen bij naam te noemen - hebben de voorbije jaren hun rechtstreekse of onrechtstreekse bijdrage geleverd aan het tot stand komen van dit proefschrift. Hierbij wil ik hen van harte bedanken. Tot een aantal personen wil ik graag een persoonlijk woord van dank richten omdat ze, op hun eigen specifieke manier, voor mij heel wat hebben betekend.

In de eerste plaats richt ik een welgemeend woord van dank aan mijn promotoren Bernard en Hans. Bernard, dank-je-wel om mij van het zinkende L&H schip te redden (alhoewel ik op dat moment nog helemaal niet wist dat het aan het zinken was) en zo mijn professionele carrière een nieuwe wending te geven. Verder wens ik je ook te bedanken om me steeds bij te staan met je gedetailleerde kennis en ervaring, om je oneindige stroom aan ideeën en om de kansen die je me hebt gegeven een stukje van de wereld te zien. Ook Hans wil ik hier bedanken voor de begeleiding en goede raad, voor de sturing van en de feedback op dit werk.

Mijn (ex-)collega's Koen, Ester en Hilde bedank ik voor het jarenlange aangename samen zitten in één bureau: het was plezant, en het wederzijdse begrip, de steun en aanmoedigingen hebben vaak heel erg deugd gedaan. Ook mijn overige collega's van de vakgroep wil ik bedanken voor alle administratieve hulp, de aangename werkomgeving en de ontspannende lunch- en koffiepauzes.

Naast alle collega's waren ook mijn vrienden en (schoon)familie erg belangrijk. Ik bedank Elke, Peter, Lotte & Robin, Tim & Els, Sven & Ciska, Fré en alle anderen die er de afgelopen jaren steeds waren als het opnieuw hoog tijd werd om over iets anders dan werk te praten. Ook Mieke & Guido, Brecht, Ann & Tom wil ik hier bedanken, een betere schoonfamilie had ik mij niet kunnen voorstellen! En natuurlijk ook

Jeroen, Ilse & Lisa, alhoewel we elkaar de laatste tijd niet meer zo vaak zien, beloof ik dat daar in de toekomst verandering in komt! Als laatste bedank ik mijn ouders, gewoon om wat ze voor mij betekenen, wie ze voor me zijn en hoe ze me al een heel leven lang steunen en motiveren in alles wat ik onderneem.

En dan rest me nog mijn twee favorieten te bedanken. Bart, ik herinner me nog goed de dag dat ik besloot mijn job bij L&H op te geven en de overstap naar de universiteit te wagen. We vroegen ons af hoe we het de komende jaren zouden klaar spelen, allebei hard werkend aan een doctoraat. We hebben samen afgezien, maar nog meer mooie momenten beleefd (hoe groot was de kans nu alweer dat we allebei op dezelfde plaats en op hetzelfde moment in Hawaï moesten zijn ?) We zijn ongeveer samen begonnen aan dit avontuur en samen ronden we het af. Eindelijk zijn we klaar met de laptops (alhoewel), laat die toekomst zonder boekjes nu maar komen!

Tenslotte wil ik Lennert, onze kleine deugniet, bedanken voor zijn stralende (ochtend)lach waarmee hij me telkens de energie geeft om de dag met hernieuwde moed en inspiratie te starten.

Gent, september 2006
Saskia Janssens

Contents

Chapter 1

Introduction

Similarity plays a fundamental role in our daily life. The importance of similarity is often underestimated, but it is clearly pointed out by researchers in psychology. Similarity serves as an organizing principle by which we classify objects, form concepts and make generalizations. Therefore, it is an important concept, not only in psychology but also in many research areas, such as biology, chemistry, information retrieval, machine learning, statistics and many more. The main goal in all these disciplines is to analyze data sets in order to find patterns and regularities that contain important knowledge about the data. Similarity of objects is one of the central concepts in data mining and knowledge discovery.

Numerous definitions of similarity measures have already been proposed in the literature. Some of them are limited in use to a specific domain, while others are widely spread and applied in about every discipline. It is worth mentioning that there does not seem to exist a strict mathematical definition of a similarity measure, at most a general understanding about the ordinal interpretation of it: the higher the value, the more similar the objects. Reflexivity is generally accepted as a basic property of a similarity measure, but this is certainly not the case for the property of symmetry.

The origin of many similarity measures can be found in the field of numerical taxonomy, where a feature-based approach is followed. Objects are represented by sets of features and in this way, reduced to $\{0,1\}$-vectors (or binary vectors). Then, similarity is based on feature commonality and difference. It is self-evident that one loses a lot of information by using $\{0,1\}$-vectors. Imagine that you would rely on black/white arguments to make choices. In most cases, a more thoughtful approach to a problem is necessary, since in the real world, many shades of gray exist between black and white. And also in many scientific domains, it is more likely that objects are transformed into real-valued vectors, rather than binary ones (for example, microarray data, document term weights, etc.). Any real-valued vector can be scaled to the unit interval, and in this way transformed into a $[0,1]$-vector. As $\{0,1\}$-vectors can be identified with ordinary sets, $[0,1]$-vectors can be represented by fuzzy sets.

Fuzzy sets were introduced by Zadeh in 1965 [95] and allow to define intermediate values between conventional evaluations like yes/no, true/false, black/white. A fuzzy set is nothing else but a mapping A from a universe X to the unit interval, where $A(x)$ denotes the degree to which x belongs to the fuzzy set A. The closer the value of $A(x)$ is to 1, the more x belongs to A. The use of a numerical scale such as the unit interval allows for a convenient representation of the gradation in membership. Therefore, fuzzy set theory provides a perfect toolbox to generalize many existing similarity measures for binary vectors into similarity measures for $[0,1]$-vectors.

When comparing two objects, one tends to use a similarity measure to express to what degree these objects are alike. Similarity, however, is only one of the conceptual ways for looking at possible relationships between objects. It is equally natural to ask for a measure expressing the degree to which one of two given objects is contained in the other, or, in other words, the degree to which one of the two is covered by the other one. In the attempt to answer this question, one is naturally led to the concept of inclusion measure. For example, defining how well a phylogenetic tree is included in the "tree of life", searching for chemical compounds or querying a document database are just some of the many applications of measuring inclusion.

Before we proceed, let me guide you through this dissertation. An overview of already existing similarity measures for binary vectors is given in Chapter 2. As binary vectors can be identified with ordinary sets, it is natural to convert these comparison measures for binary vectors into measures for ordinary sets. The latter are then based on the cardinalities of the sets involved. Furthermore, we introduce some already existing parametric families of similarity and inclusion measures for ordinary sets and introduce two new parametric families as well. Since similarity and inclusion measures are defined as fuzzy relations and since T-transitivity (with T an arbitrary t-norm) is one the most important properties that can be attributed to fuzzy relations, it is worth to investigate this property. Moreover, since $T_{\mathbf{L}}$- and $T_{\mathbf{P}}$-transitivity are highly related to the concept of a pseudometric, we identify the $T_{\mathbf{L}}$- and $T_{\mathbf{P}}$-transitive members of our families. Conversely, $T_{\mathbf{M}}$-transitivity is related to the concept of an ultrametric, but we conclude that none of two introduced families have $T_{\mathbf{M}}$-transitive members.

In Chapter 3, an introduction to fuzzy set theory is given. Throughout the sequel of this work, the sigma count is used as the cardinality of a fuzzy set and we will refer to it as the basic scalar cardinality. Furthermore, the validity of generalizations of identities on crisp cardinalities using t-norms and t-conorms (such as the valuation property and the fuzzification of the cardinality of the (symmetric) difference of two fuzzy sets) are investigated on a one-by-one basis. The reader who is already familiar with concepts such as triangular norm, (quasi)-copula, sigma count and the fuzzification of intersection and union of two fuzzy sets can easily skip this chapter.

The Bell inequalities are introduced in Chapter 4 and are rewritten in the context of basic scalar cardinalities. We prove that some inequalities are fulfilled for (quasi-)copulas. Moreover, considering the Frank t-norm family and the major parametric t-norm families, we identify all parameter values such that each of the Bell-type inequalities is fulfilled. A major contribution of this chapter is that ordinal sums preserve the Bell-type inequalities.

In Chapter 5, we demonstrate that the Bell-type inequalities are of particular interest in the context of cardinalities of fuzzy sets. Moreover, the results on the fuzzified Bell-type inequalities can be exploited to develop a framework in which the validity of more general inequalities on fuzzy cardinalities can be checked easily.

The parametric family of similarity measures for ordinary sets, which was introduced in Chapter 2 is fuzzified in Chapter 6. Next to an overview of already existing fuzzy similarity measures, we also investigate the T-transitive members (with T one of the three basic t-norms $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ or $T_{\mathbf{L}}$) of this parametric family of fuzzy similarity measures. On the other hand, the parametric family of inclusion measures for ordinary sets, also introduced in Chapter 2, is fuzzified in Chapter 7. Also for this family of fuzzy inclusion measures, we identify its T-transitive members (again, with T one of the three basic t-norms $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ or $T_{\mathbf{L}}$).

How these parametric families of fuzzy similarity or inclusion measures can be used in real-world applications is shown in Chapter 8.

A Dutch summary of this work can be found in Chapter 9.

Research performed within the context of this PhD thesis resulted in several publications in peer reviewed high-impact international journals [22, 44, 47, 48] and in international conference proceedings [26, 42, 43, 45, 46].

Chapter 2

Comparison measures for ordinary sets

2.1 Introduction

In daily life, we are often confronted with situations where we need to distinguish between several objects. Therefore, similarity tools are of great importance to obtain a degree of resemblance between two or more objects. Also in many scientific domains, such as chemistry, biology, information retrieval and many more, we come across similarity problems, so the need arises for appropriate similarity measures. Similarity, however, is only one of the conceptual ways for looking at possible relationships between objects. With a growing number of data available, it is necessary to organize this data into files or tables and in this way, to construct databases. Then, the need to extract knowledge from these databases is essential. Therefore, among other things, methods for searching a database are unavoidable. With this in mind, one takes a query object and searches for a data object, stored in a database, which has the highest match with the query object. In the attempt to answer this question, one is naturally led to the concept of inclusion.

The simplest way to compare two objects is to provide an appropriate set of features typical for those two objects and to construct for each object a binary vector encoding the presence (1) or absence (0) of each of these features. The degree of similarity (or inclusion) of two objects is then often expressed in terms of the cardinalities of the latter sets. In this way, one can simply compare binary vectors, rather than comparing the objects themselves.

There is a vast amount of papers concerning similarity measures

		vector B	
		1	0
vector A	1	a	b
	0	c	d

Table 2.1: Outcome possibilities when comparing two binary vectors.

available in the literature, nevertheless the contribution to inclusion measures is rather poor. What strikes one most is that no definition of similarity is employed, let alone a definition of inclusion. In literature, a similarity or inclusion measure is just a function that assigns a non-negative, real number, defining a notion of resemblance. Also the term "similarity" is rather uncommon, usually one talks about association coefficient, proximity, resemblance or sometimes dissimilarity. "Inclusion" is an established expression, although one also stumbles across the term "subsethood".

In this chapter, we give an overview of commonly used similarity and inclusion measures for binary vectors. As binary vectors can be represented by ordinary sets, we translate the comparison measures for binary vectors into comparison measures for ordinary sets. Furthermore, we give an overview of already existing parametrized families of similarity measures and introduce a new parametrized family for both similarity and inclusion measures. As T-transitivity is highly related to the concept of a metric, we investigate the T-transitivity properties, with T one of the three basic t-norms, for the two families mentioned above.

2.2 Commonly used similarity measures for binary vectors

For each object, a binary vector is constructed by selecting an appropriate set of features encoding the presence or absence of each of these features. Then, the comparison of those vectors instead of the objects themselves is used a lot in practice and, not surprisingly, many similarity measures based on the presence or absence of common features already exist. An overview of such similarity measures can be found, for instance, in Sneath and Sokal [79]. The most popular similarity measure encountered in about every discipline is still the Jaccard coefficient [41].

As usual, the similarity measures used to compare binary vectors are based on Table 2.1 where a denotes the number of features that are common to both objects, where b and c denote the number of features that are present in only one object and where d denotes the number of features that both objects lack.

Example 2.1 Suppose we have three objects represented by a binary vector.

Object 1	1	0	0	0	1	1	0	0	1	0
Object 2	0	0	0	0	1	0	0	1	1	0
Object 3	0	0	0	0	0	0	0	1	0	0

Computing the similarity using the Jaccard coefficient ($J(A,B) = \frac{a}{a+b+c}$, with a, b and c defined as in Table 2.1), we obtain the following:

$$\begin{aligned} J(\text{Object }1, \text{Object }2) &= \frac{2}{2+1+2} = 0.4\,, \\ J(\text{Object }1, \text{Object }3) &= \frac{0}{3+1} = 0\,, \\ J(\text{Object }2, \text{Object }3) &= \frac{1}{1+2} \approx 0.33\,. \end{aligned}$$

According to Gower and Legendre [39], one of the criteria to choose an appropriate similarity measure for a certain problem is whether one should include conjoint absences or not. With the choice of a suitable measure, often the term d leads to a discussion. In some circumstances, it would seem ridiculous to compare two species (or more general two objects) on the basis of the features they both lack, whereas in other situations it would seem improper to neglect these conjoint absences. For example, the absence of wings when observed among a group of distantly related organisms (such as camel, louse and nematode) would be an absurd indication of similarity [79]. On the other hand, when pharmaceutical drugs are evaluated in a clinical trial with respect to the response of a number of patients treated with the drug (where 1 is used if a drug has a positive effect and 0 if it has a negative effect) it is important that both the positive and negative matches should contribute to similarity [72]. Some authors (e.g. Clifford and Stephenson [17]) classify the commonly used similarity measures into two groups: measures that exclude occurrences of negative matches and the ones that include these occurrences. Others (e.g. Sneath and Sokal [79]) use a classification based on whether the numerator contains occurrences of negative matches or not. In Section 2.4.3, we present a parametrized family of similarity measures with the possibility to attribute a certain weight to the contribution of negative matches.

Commonly used measures that include occurrences of negative matches are for instance the following measures: the simple matching coefficient [78], the Russel–Rao coefficient [70], the Rogers–Tanimoto coefficient [69], the Hamann coefficient [40], and the Sokal–Sneath coefficients [79]. The Jaccard coefficient [41], the Dice coefficient [28], the cosine coefficient [62], the Fager–McGowan coefficient [32] and the Kulczynski coefficient [56] on the other hand belong to the category of measures that exclude occurrences of negative matches. An overview of these coefficients can be found in Tables 2.2 and 2.3. There are so many similarity measures for binary vectors in the literature available that any attempt at an exhaustive catalog of them (with details about use and origin) would require many pages and it is doubtful whether such a dry enumeration would be of additional value to the reader. Therefore, only those coefficients that have been used extensively in the literature (like the Jaccard coefficient, the Dice coefficient, the simple matching coefficient and the cosine coefficient) are discussed in more detail.

Coefficient	$S(A,B) =$	Range	θ	ϕ
Jaccard [41]	$\frac{a}{a+b+c}$	$[0,1]$	–	1
Simple matching [78]	$\frac{a+d}{a+b+c+d}$	$[0,1]$	1	–
Dice [28]	$\frac{2a}{2a+b+c}$	$[0,1]$	–	$\frac{1}{2}$
Rogers and Tanimoto [69]	$\frac{a+d}{a+2(b+c)+d}$	$[0,1]$	2	–
Sokal and Sneath 1 [79]	$\frac{a}{a+2(b+c)}$	$[0,1]$	–	2
Sokal and Sneath 2 [79]	$\frac{2a+2d}{2a+b+c+2d}$	$[0,1]$	$\frac{1}{2}$	–

Table 2.2: Overview of similarity measures - part 1.

Coefficient	$S(A,B) =$	Range
Hamann [40]	$\frac{a+d-b-c}{a+b+c+d}$	$[-1,1]$
Russell and Rao [70]	$\frac{a}{a+b+c+d}$	$[0,1]$
Kulczynski 1 [56]	$\frac{a}{b+c}$	$[0,+\infty]$
Kulczynski 2 [56]	$\frac{a}{2}\left(\frac{1}{a+b}+\frac{1}{a+c}\right)$	$[0,1]$
Ochiai, Cosine [62]	$\frac{a}{\sqrt{(a+b)(a+c)}}$	$[0,1]$
Fager and McGowan [32]	$\frac{a}{\sqrt{(a+b)(a+c)}} - \frac{1}{2\sqrt{a+b}}$	$[0,1]$
Pearson [72]	$\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$	$[-1,1]$
Simpson, Overlap [75]	$\frac{a}{a+\min(b,c)}$	$[0,1]$
Forbes [28]	$\frac{a(a+b+c+d)}{(a+b)(a+c)}$	$[0,+\infty]$
Braun-Blanquet [11]	$\frac{a}{a+\max(b,c)}$	$[0,1]$
Baroni-Urbani and Buser [3]	$\frac{a+\sqrt{ad}}{a+b+c+\sqrt{ad}}$	$[0,1]$
Yule [94]	$\frac{ad-bc}{ad+bc}$	$[-1,1]$

Table 2.3: Overview of similarity measures - part 2.

Jaccard coefficient. The Jaccard coefficient [41] is the most popular measure in about every discipline. It is a measure that omits occurrences of negative matches and reads as the number of common features in both objects over the number of all features present in either one of them.This measure is used in any field where similarity between objects is involved, notwithstanding the fact that in some disciplines the term Tanimoto coefficient is used more frequently than the original Jaccard coefficient (e.g. in the field of virtual screening of chemical structure databases [88]). Note that the first coefficient of Sokal and Sneath resembles the Jaccard coefficient except that mismatches carry a double weight.

Simple matching coefficient. The simple matching coefficient is one of the simplest and oldest measures, which was introduced by Sokal and Michener [78], to deal with the evaluation of taxonomic relationships, and used repeatedly ever since in many disciplines. This coefficient is very similar to the one of Jaccard except that it includes occurrences of negative matches for measuring similarity.

Note that this coefficient corresponds to different expressions depending on the domain of practice. In information retrieval models, the similarity between two documents is often based on the number of keywords (terms). Then, one often uses as the simple matching coefficient just the number of keywords common to a pair of documents [82, 86], i.e. $S(A, B) = a$[1]. The origin of this expression can be found in the so-called "vector space model of information retrieval": documents are represented by (not necessarily) binary vectors of (weights of) terms. Taking the inner product of two vectors renders the similarity between them. When documents are represented by binary vectors of terms, it is easy to see that the inner product of two vectors reduces to the number of terms common to both documents.

This has the disadvantage that the similarity is not normalized, i.e. it takes no account of the numbers of terms in each of the documents, which is a severe limitation and thus most coefficients that have been used try to normalize this definition of the simple matching coefficient in some way [82]. Note that in this way, the Jaccard, Dice, cosine and overlap coefficients are a kind of normalization of the simple matching coefficient.

In other domains (like information theory and coding theory) the simple matching coefficient is often referred to as the Hamming distance.

In what follows, we always have in mind the definition of Sokal and Michener when talking about the simple matching coefficient. Note that as well as the coefficient of Rogers and Tanimoto, the second coefficient of Sokal and Sneath as the Hamann coefficient resemble the simple matching coefficient.

Dice coefficient. In many ecological studies there is a need to express the degree to which two different species are associated in nature. In 1945, Dice proposed his association coefficient [28], because of the lack of a measure that was common use among ecologists. First, he in-

[1] Note that this definition of the simple matching coefficient is related to the Hamming distance. The Hamming distance between two strings is defined as the number of positions for which the corresponding symbols are different.

troduced an association index which may differ depending on which species is used as the basis of comparison. Let $a + b$ be the number of samples in which species A occurs and a be the number of samples in which species A and B occur together, then the association index of species B with species A (denoted by B/A) is defined by $\frac{a}{a+b}$. Analogously, the association index of species A with species B (denoted by A/B) is defined by $\frac{a}{a+c}$, where $a + c$ denotes the number of samples in which species B occurs. Nevertheless, Dice noticed that in some ecological studies it is desirable to have a measure at one's disposal that does not change depending on which species is used as a base. Such a measure, which he called the coincidence index, has a value intermediate between the association indices B/A and A/B and is defined as:

$$S(A, B) = \frac{2a}{2a + b + c} \,.$$

Further on, we will refer to this similarity measure as the Dice coefficient. Ecologists often use different names to describe the same index. For example, the Czekanowski coefficient, the Sørensen coefficient and the Dice coefficient are all synonyms [17, p. 55].

Note that the Dice coefficient is similar to the Jaccard coefficient but gives twice the weight to agreements.

Ochiai/Cosine coefficient. The Ochiai coefficient is also known as the cosine coefficient. This association coefficient is commonly used in document clustering techniques (note that it is some normalization of the simple matching coefficient as defined in document clustering).

The cosine of the angle α between two vectors $V = (v_1, \ldots, v_n)$ and $W = (w_1, \ldots, w_n)$ in the n-dimensional space is equal to

$$\cos\alpha = \frac{\sum_{i=1}^{n} v_i w_i}{\sqrt{\sum_{i=1}^{n} v_i \sum_{i=1}^{n} w_i}} \,.$$

It is easy to see that the similarity of two binary vectors can be calculated using the Ochiai/cosine coefficient as indicated in Table 2.3.

Gower and Legendre [39] have introduced two one-parameter families S_θ and T_ϕ of similarity measures for binary vectors, with θ and ϕ positive reals. Using the notations introduced above, these families can be written as:

$$\begin{aligned} S_\theta(A, B) &= \frac{a + d}{a + d + \theta(b + c)} \,, \\ T_\phi(A, B) &= \frac{a}{a + \phi(b + c)} \,. \end{aligned}$$

Some of the measures in Table 2.2 belong to one of these families, as indicated in the table.

2.3 From binary vectors to ordinary sets

Each binary vector $(a_1, \ldots, a_n) \in \{0,1\}^n$ can be represented by an ordinary subset A in a finite universe $X = \{x_1, x_2, \ldots, x_n\}$ of cardinality n (the feature space) in the following way:

$$a_i = 1 \Leftrightarrow x_i \in A\,.$$

When a bit i is set on in a binary vector, then the corresponding feature x_i belongs to a subset A of X and vice versa. The degree of similarity of two objects (or two binary vectors) is then expressed in terms of the cardinalities of the latter sets.

If A and B are the set representations of two binary vectors a and b, then the number of features that are common to both vectors are represented by the cardinality of $A \cap B$, denoted $|A \cap B|$. The number that are present in a, but not in b or vice versa is identical to $|A \setminus B|$ or $|B \setminus A|$, respectively. Finally, $|(A \cup B)^c|$ equals the number of features which both objects lack.

In this way, Table 2.1 with the outcome possibilities when comparing two binary vectors can be translated into Table 2.4.

		set B	
		$x_i \in B$	$x_i \notin B$
set A	$x_i \in A$	$\lvert A \cap B\rvert$	$\lvert A \setminus B\rvert$
	$x_i \notin A$	$\lvert B \setminus A\rvert$	$\lvert (A \cup B)^c\rvert$

Table 2.4: Outcome possibilities when comparing two ordinary sets.

Then, all similarity measures for binary vectors listed in Tables 2.2 and 2.3 can easily be translated into similarity measures for ordinary sets. However, since we restrict ourselves to the measures from Table 2.2 in the remainder of this work, we only translate these measures into measures for ordinary sets and recapitulate them in Table 2.5. Note that $|A \triangle B| = |A \setminus B| + |B \setminus A|$.

At this moment, we are able to give a formal definition of a similarity measure for ordinary sets. Therefore, we need the concept of a fuzzy set and the related concept of a binary fuzzy relation.

Measure	Expression	θ	ϕ	Metric
Jaccard	$\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$	–	1	yes
Simple Matching	$1-\frac{\lvert A\triangle B\rvert}{n}$	1	–	yes
Dice	$\frac{2\lvert A\cap B\rvert}{\lvert A\triangle B\rvert+2\lvert A\cap B\rvert}$	–	$\frac{1}{2}$	no
Rogers and Tanimoto	$\frac{n-\lvert A\triangle B\rvert}{n+\lvert A\triangle B\rvert}$	2	–	yes
Sokal and Sneath 1	$\frac{\lvert A\cap B\rvert}{\lvert A\cap B\rvert+2\lvert A\triangle B\rvert}$	–	2	yes
Sokal and Sneath 2	$1-\frac{\lvert A\triangle B\rvert}{2n-\lvert A\triangle B\rvert}$	$\frac{1}{2}$	–	no

Table 2.5: Some well-known cardinality-based similarity measures.

Fuzzy sets are a generalization of classical sets and were introduced by Zadeh in 1965 as a mathematical means to represent vagueness in everyday life [95]. They were specifically designed to provide formalized tools for dealing with the imprecision intrinsic to many problems.

A fuzzy set A in a universe X is characterized by an $X \to [0,1]$ mapping where $A(x)$ is interpreted as the degree of membership of element x in the fuzzy set A, for any $x \in X$. The value 0 means that the element is not included in the given set, the value 1 describes a fully included member (this behavior corresponds to ordinary sets), while values in between are used to represent intermediate degrees of membership. The family of all fuzzy sets in X will be denoted by $\mathcal{F}(X)$.

Definition 2.1 *A binary fuzzy relation R is a mapping $R : X^2 \to [0,1]$ where $R(x,y)$ denotes the degree to which x is related to y.*

Note that any fuzzy relation R is a fuzzy subset of $X \times X$, i.e. $R \in \mathcal{F}(X \times X)$. As we are dealing with binary fuzzy relations only, we will drop the adjective binary further on in this text. A fuzzy relation R on a finite universe $X = \{x_1, x_2, \ldots, x_n\}$ can also be represented by means of the matrix A_R, with elements $a_{ij} = R(x_i, x_j)$ for any $i, j = 1, \ldots, n$. For $0 \leq \alpha \leq 1$, the α-cut R_α of a fuzzy relation R on X is the crisp relation on X defined by:

$$(x,y) \in R_\alpha \Leftrightarrow R(x,y) \geq \alpha\,.$$

Example 2.2 Let $X = \{\text{Amsterdam, Brussels, Paris}\}$. Let R be a fuzzy relation that represents the notion of "far". Then R can be defined as follows:

$$R(\text{Amsterdam}, \text{Brussels}) = 0.3\,,$$
$$R(\text{Amsterdam}, \text{Paris}) = 0.8\,,$$
$$R(\text{Brussels}, \text{Paris}) = 0.5\,.$$

Equivalently, R can be represented by the 3×3 matrix A_R :

$$A_R = \begin{bmatrix} 1 & 0.3 & 0.8 \\ 0.3 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{bmatrix}.$$

Definition 2.2 *A similarity measure for ordinary sets is a reflexive, symmetrical fuzzy relation on the power set $\mathcal{P}(X) = \{0,1\}^X$.*

In this way, a similarity measure S fulfills the following properties for any subsets A and B of X:

(i) $S(A,B) \in [0,1]\,,$

(ii) Reflexivity: $S(A,A) = 1\,,$

(iii) Symmetry: $S(A,B) = S(B,A)\,.$

Typically, judgements of similarity are assumed to be equivalent to judgements of dissimilarity and vice versa. The function relating these concepts is generally assumed to be an inverse relation. For example, if the similarity between objects A and B equals $S(A,B)$, then the dissimilarity between A and B is equal to $D(A,B) = 1 - S(A,B)$. In this way, all similarity measures from Table 2.2 and some of Table 2.3 (only those of which the range is equal to the unit interval) can be converted to a dissimilarity measure. A dissimilarity measure represents the discrepancy between two objects and intuitively, it is strongly related to a measure of distance. Therefore, it is desirable that a dissimilarity measure observes the properties of a metric.

Definition 2.3 *A mapping $d : X \times X \rightarrow \mathbb{R}^+$ is a metric if for any $(x,y,z) \in X^3$ holds that:*

(i) $d(x,y) = 0 \Leftrightarrow x = y$

(ii) *Symmetry:* $d(x,y) = d(y,x)$

(iii) *Triangle inequality:* $d(x, z) \leq d(x, y) + d(y, z)$

In [39], it was already proven that the families $1-S_\theta$ and $1-T_\phi$ fulfill the properties of a metric for $\theta \geq 1$ and $\phi \geq 1$, respectively. This means that the Jaccard coefficient, the simple matching coefficient, the coefficient of Rogers and Tanimoto and the first coefficient of Sokal and Sneath are metric (see Table 2.5). Note that Gower and Legendre also studied the metric properties of $\sqrt{1-S_\theta}$ and $\sqrt{1-T_\phi}$, but since this is of no interest in the remainder of this thesis, we do not pay attention to it.

If instead of (i) a dissimilarity measure d satisfies for any $(x, y) \in X^2$

(i′) $d(x, y) = 0 \Leftarrow x = y$

then d is called a pseudo-metric. Therefore, every metric is also a pseudo-metric. When the third property is strengthened to

(iii′) $d(x, z) \leq \max(d(x, y), d(y, z))$

then d is called an ultrametric. The importance of metrics, pseudo-metrics and ultrametrics will be clarified in Section 2.5.

In the same way as we defined a similarity measure, we can give a definition of an inclusion measure for ordinary sets as follows.

Definition 2.4 *An inclusion measure for ordinary sets is a binary fuzzy relation I on the power set $\mathcal{P}(X) = \{0, 1\}^X$ satisfying:*

$$A \subseteq B \Rightarrow I(A, B) = 1\,. \tag{2.1}$$

Note that the above condition only involves subsets of X that are contained into one another. It expresses that any subset A of a given set B is equally well included in that set as the set into itself and this with maximum degree of inclusion 1, or equivalently, that a set covers all its subsets to the same degree 1 as it covers itself. Condition (2.1) implies the property of reflexivity. Indeed, by setting B equal to A in condition (2.1), one obtains that $I(A, A) = 1$, for any $A \in \mathcal{P}(X)$.

As already indicated, there are only a few inclusion measures based on the cardinality of ordinary sets known in the literature. Kuncheva [57] defines an inclusion measure for two ordinary sets A and B as follows: $I(A, B) = \frac{|A \cap B|}{|A|}$. Willmott [90] also introduced a cardinality-based inclusion measure : $I(A, B) = \frac{|B|}{|A \cup B|}$. Furthermore, in the paper of De Baets, De Meyer and Naessens [21] several cardinality-based inclusion measures are discussed.

2.4 Parametric families of comparison measures

2.4.1 Tversky's contrast model

Using set theory, Tversky [81] defined a similarity measure as a feature-matching process. It produces a similarity value that is not only the result of common features, but also the result of the differences between two objects. The matching process is defined by $A \cap B$, $A \setminus B$ and $B \setminus A$. Tversky's contrast model defines the similarity between two objects A and B, as follows:

$$S(A,B) = \theta f(A \cap B) - \alpha f(A \setminus B) - \beta f(B \setminus A)\,,$$

with θ, α, $\beta \geq 0$ and f an appropriate interval scale satisfying $f(X \cup Y) = f(X) + f(Y)$ whenever X and Y are disjoint. The parameters θ, α and β refer to the weights for common and different features between the two objects.

Using the cardinality of an ordinary set as an appropriate interval scale, we can translate Tversky's contrast model into the following family of cardinality-based similarity measures:

$$S(A,B) = \theta|A \cap B| - \alpha|A \setminus B| - \beta|B \setminus A|\,.$$

A disadvantage of Tversky's contrast model is that this similarity measure is not bounded by 1. A matching function that normalizes the value of similarity is the ratio model [81]

$$S(A,B) = \frac{|A \cap B|}{|A \cap B| + \alpha|A \setminus B| + \beta|B \setminus A|}\,, \tag{2.2}$$

with $\alpha, \beta \geq 0$. Note that this family of similarity measures is reflexive, but not symmetric. The ratio model generalizes several similarity measures proposed in the literature. For example, by setting $\alpha = \beta = 1$, we obtain the Jaccard coefficient, substituting $\alpha = \beta = \frac{1}{2}$ yields the Dice coefficient, while substituting $\alpha = \beta = 2$ yields the first coefficient of Sneath and Sokal. By setting $\alpha = 1$ and $\beta = 0$ we retrieve the Braun–Blanquet coefficient if $|B \setminus A| \leq |A \setminus B|$. In case $|B \setminus A| > |A \setminus B|$ we obtain the overlap coefficient. Remark that this coefficient is also equal to the inclusion measure defined by Kuncheva [57].

Remark that the weights allow for the definition of an asymmetric measure for similarity. Indeed, symmetry holds whenever the objects are equal in measure (i.e. $f(A) = f(B)$) or the task is non-directional (i.e. $\alpha = \beta$). To interpret the latter condition, compare the following two forms:

1. assess the degree to which object A and object B are similar to each other (a non-directional task, i.e. $\alpha = \beta$),

2. assess the degree to which object A is similar to object B (a directional task, i.e. $\alpha \neq \beta$).

If $S(A, B)$ is interpreted as the degree to which A is similar to B, then A is the subject of comparison and B is the referent.

The directionality and asymmetry of a similarity measurement task are particularly noticeable in word associations. We say "the son resembles the father" rather than "the father resembles the son". But also in scientific domains, such as chemical similarity, asymmetric similarity is often used. For example, clustering chemical compounds corresponds to a non-directional task, while querying a database corresponds to a directional task [34, 74, 87, 88]. In this respect, inclusion measures can be seen as asymmetric similarity measures.

A thorough (psychological) analysis whether a similarity measure should be symmetric or not can be found in [81].

2.4.2 The families of Gower and Legendre

Gower and Legendre [39] have introduced two one-parameter families S_θ and T_ϕ of similarity measures for binary data, with θ and ϕ positive reals (which were already mentioned in Section 2.2). Using the notations introduced above, these families can be easily translated into families of cardinality-based similarity measures:

$$\begin{aligned} S_\theta(A,B) &= \frac{|A \cap B| + |(A \cup B)^c|}{\theta|A \triangle B| + |A \cap B| + |(A \cup B)^c|}\,, \\ T_\phi(A,B) &= \frac{|A \cap B|}{\phi|A \triangle B| + |A \cap B|}\,, \end{aligned}$$

with $|A \triangle B| = |A \setminus B| + |B \setminus A|$. Note that these similarity measures are reflexive and symmetric for any $A, B \in \mathcal{P}(X)$.

Each of the measures in Table 2.5 belongs to one of these families, as indicated in the table. Note that family T_ϕ corresponds to the symmetric part ($\alpha = \beta = \phi$) of family (2.2) introduced by Tversky [81].

2.4.3 A parametric family of cardinality-based similarity measures

De Baets, De Meyer and Naessens [20] have dealt with the systematic construction of cardinality-based similarity measures for comparing ordinary subsets of a finite universe $X = \{x_1, x_2, \ldots, x_n\}$. More specifically, attention was focused on a class of $[0,1]$-valued similarity measures that are rational expressions in the cardinalities of the sets involved:

$$S(A,B) = \frac{x\,\alpha_{A,B} + t\,\omega_{A,B} + y\,\delta_{A,B} + z\,\nu_{A,B}}{x'\,\alpha_{A,B} + t'\,\omega_{A,B} + y'\,\delta_{A,B} + z'\,\nu_{A,B}}\,, \tag{2.3}$$

with $A, B \in \mathcal{P}(X)$,

$$\begin{aligned} \alpha_{A,B} &= \min(|A \setminus B|, |B \setminus A|)\,, \\ \omega_{A,B} &= \max(|A \setminus B|, |B \setminus A|)\,, \\ \delta_{A,B} &= |A \cap B|\,, \\ \nu_{A,B} &= |(A \cup B)^c|\,, \end{aligned}$$

and $x, y, z, t, x', y', z', t' \in \{0,1\}$. Note that these similarity measures are symmetric for any $A, B \in \mathcal{P}(X)$. Reflexive similarity measures are characterized by $y = y'$ and $z = z'$. In this thesis, we restrict our attention to the (still large) subclass obtained by putting $t = x$ and $t' = x'$:

$$S(A,B) = \frac{x\,\triangle_{A,B} + y\,\delta_{A,B} + z\,\nu_{A,B}}{x'\,\triangle_{A,B} + y\,\delta_{A,B} + z\,\nu_{A,B}}\,, \tag{2.4}$$

with $\triangle_{A,B} = |A \triangle B|$. On the other hand, the parameters x, y, z and x' are considered to be positive reals. In this way, the class of cardinality-based similarity measures (2.3) is enlarged such that a wide spectrum of similarity measures can be provided to the reader.

In order to guarantee that $S(A,B) \in [0,1]$, we need to impose the restriction $0 \leq x \leq x'$. Since the case $x = x'$ leads to trivial measures taking value 1 only, we consider from here on $0 \leq x < x'$.

The similarity measures gathered in Table 2.5 all belong to our family of similarity measures (2.4); the corresponding parameter values are indicated in Table 2.6.

It is easy to see that families S_θ and T_ϕ introduced by Gower and Legendre belong to the family of similarity measures (2.4). Indeed, S_θ can be obtained by setting $x = 0$, $y = z = 1$ and $x' = \theta$, while substituting $x = z = 0$, $y = 1$ and $x' = \phi$ leads to T_ϕ.

Measure	Expression	x	x'	y	z	T
Jaccard	$\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$	0	1	1	0	$T_{\mathbf{L}}$
Simple Matching	$1-\frac{\lvert A\triangle B\rvert}{n}$	0	1	1	1	$T_{\mathbf{L}}$
Dice	$\frac{2\lvert A\cap B\rvert}{\lvert A\triangle B\rvert+2\lvert A\cap B\rvert}$	0	1	2	0	–
Rogers and Tanimoto	$\frac{n-\lvert A\triangle B\rvert}{n+\lvert A\triangle B\rvert}$	0	2	1	1	$T_{\mathbf{L}}$
Sokal and Sneath 1	$\frac{\lvert A\cap B\rvert}{\lvert A\cap B\rvert+2\lvert A\triangle B\rvert}$	0	2	1	0	$T_{\mathbf{L}}$
Sokal and Sneath 2	$1-\frac{\lvert A\triangle B\rvert}{2n-\lvert A\triangle B\rvert}$	0	1	2	2	–

Table 2.6: Some members of family (2.4).

2.4.4 A parametric family of cardinality-based inclusion measures

In [21], a systematic way of generating inclusion measures for ordinary sets on a finite universe $X=\{x_1,\ldots,x_n\}$ in the form of a rational expression solely based on cardinalities of the sets involved was presented:

$$I(A,B)=\frac{x\,\chi_{A,B}+t\,\chi_{B,A}+y\,\delta_{A,B}+z\,\nu_{A,B}}{x'\,\chi_{A,B}+t'\,\chi_{B,A}+y'\,\delta_{A,B}+z'\,\nu_{A,B}}\,,$$

with $\chi_{A,B}=|A\setminus B|$, $\chi_{B,A}=|B\setminus A|$ and $x,y,z,t,x',y',z',t'\in\{0,1\}$. The inclusion measures satisfying (2.1) are then characterized by $y=y'$, $z=z'$ and $t=t'$.

In analogy to the class of reflexive similarity measures we consider the following subclass, by putting $t=x'$:

$$I(A,B)=\frac{x\,\chi_{A,B}+x'\,\chi_{B,A}+y\,\delta_{A,B}+z\,\nu_{A,B}}{x'\;\triangle_{A,B}+y\,\delta_{A,B}+z\,\nu_{A,B}}\,. \tag{2.5}$$

The parameters x, y, z and x' are considered to be positive reals. In order to guarantee that $I(A,B)\in[0,1]$, we need to impose the following restriction: $x\le x'$. Since the case $x=x'$ leads to trivial measures taking value 1 only, we consider from here on $0\le x<x'$. Note that in this case also the converse implication in (2.1) holds.

Some inclusion measures known in the literature are a member of this parametric family [21, 90]. They are summarized in Table 2.7. Note that the inclusion measure given by Kuncheva [57] is not a member of family (2.5).

Measure	Expression	x	x'	y	z	T
I_1	$\frac{\lvert B\setminus A\rvert}{\lvert A\triangle B\rvert}$	0	1	0	0	$T_{\mathbf{L}}$
I_2	$\frac{\lvert A^c\rvert}{\lvert (A\cap B)^c\rvert}$	0	1	0	1	$T_{\mathbf{P}}$
I_3	$\frac{\lvert B\rvert}{\lvert A\cup B\rvert}$	0	1	1	0	$T_{\mathbf{P}}$
I_4	$\frac{\lvert (A\setminus B)^c\rvert}{n}$	0	1	1	1	$T_{\mathbf{L}}$
I_5	$1-\frac{2\lvert A\setminus B\rvert}{n+\lvert A\triangle B\rvert}$	0	2	1	1	$T_{\mathbf{L}}$

Table 2.7: Some members of family (2.5).

2.5 T-transitivity versus pseudo-metrics

The rational similarity and inclusion measures introduced in [20] and [21], respectively, have also been investigated for properties such as monotonicity and T-transitivity, with T a triangular norm (or t-norm for short).

Definition 2.5 *[52] A binary operation $T : [0,1]^2 \to [0,1]$ is called a t-norm if it satisfies:*

(i) *Neutral element* 1*:* $(\forall x \in [0,1])(T(x,1) = x)$.

(ii) *Monotonicity: T is increasing in each variable.*

(iii) *Commutativity:* $(\forall (x,y) \in [0,1]^2)(T(x,y) = T(y,x))$.

(iv) *Associativity:* $(\forall (x,y,z) \in [0,1]^3)(T(x,T(y,z)) = T(T(x,y),z))$.

Although originating from the field of probabilistic metric spaces, t-norms are nowadays mainly popular as model for many-valued conjunction in fuzzy logic and for defining the intersection of fuzzy sets in a pointwise manner.

The three basic continuous t-norms are the minimum operator $T_{\mathbf{M}}$, the algebraic product $T_{\mathbf{P}}$ and the Łukasiewicz t-norm $T_{\mathbf{L}}$ defined by

$$\begin{aligned} T_{\mathbf{M}}(x,y) &= \min(x,y)\,, \\ T_{\mathbf{P}}(x,y) &= x \cdot y\,, \\ T_{\mathbf{L}}(x,y) &= \max(x+y-1,0)\,. \end{aligned}$$

Another basic t-norm (which is not continuous) is the drastic product $T_{\mathbf{D}}$ defined by

$$T_{\mathbf{D}}(x,y) = \begin{cases} 0 & \text{, if } (x,y) \in [0,1[^2\,, \\ \min(x,y) & \text{, otherwise}\,. \end{cases}$$

These four t-norms can be ordered as follows: $T_{\mathbf{D}} < T_{\mathbf{L}} < T_{\mathbf{P}} < T_{\mathbf{M}}$. The three continuous t-norms can be considered as prototypical cases, since any other continuous t-norm can be built starting from these basic t-norms. Every continuous t-norm can be expressed as an ordinal sum of them. Ordinal sums are defined as follows:

Definition 2.6 [52] *Let $(T_\alpha)_{\alpha \in A}$ be a family of t-norms and $(]a_\alpha, e_\alpha[)_{\alpha \in A}$ be a family of non-empty, pairwise disjoint open subintervals of $[0,1]$. The t-norm T defined by*

$$T(x,y) = \begin{cases} a_\alpha + (e_\alpha - a_\alpha) T_\alpha\left(\frac{x-a_\alpha}{e_\alpha - a_\alpha}, \frac{y-a_\alpha}{e_\alpha - a_\alpha}\right) & \text{, if } (x,y) \in [a_\alpha, e_\alpha]^2\,, \\ \min(x,y) & \text{, otherwise}\,, \end{cases}$$

is called the ordinal sum of the summands $\langle a_\alpha, e_\alpha, T_\alpha \rangle$, and we write

$$T = (\langle a_\alpha, e_\alpha, T_\alpha \rangle)_{\alpha \in A}\,.$$

Example 2.3 The ordinal sum

$$T = (\langle 0.1, 0.3, T_{\mathbf{P}} \rangle\,, \langle 0.3, 0.6, T_{\mathbf{L}} \rangle\,, \langle 0.8, 1, T_{\mathbf{L}} \rangle)$$

is given by

$$T(x,y) = \begin{cases} 0.1 + 5(x-0.1)(y-0.1) & \text{, if } (x,y) \in [0.1, 0.3]^2\,, \\ 0.3 + \max(x+y-0.9, 0) & \text{, if } (x,y) \in [0.3, 0.6]^2\,, \\ 0.8 + \max(x+y-1.8, 0) & \text{, if } (x,y) \in [0.8, 1]^2\,, \\ \min(x,y) & \text{, otherwise}\,. \end{cases}$$

The important parts of the domain of

$$T = (\langle 0.1, 0.3, T_{\mathbf{P}} \rangle\,, \langle 0.3, 0.6, T_{\mathbf{L}} \rangle\,, \langle 0.8, 1, T_{\mathbf{L}} \rangle)$$

and its $3D$ plot are depicted in Figure 2.1.

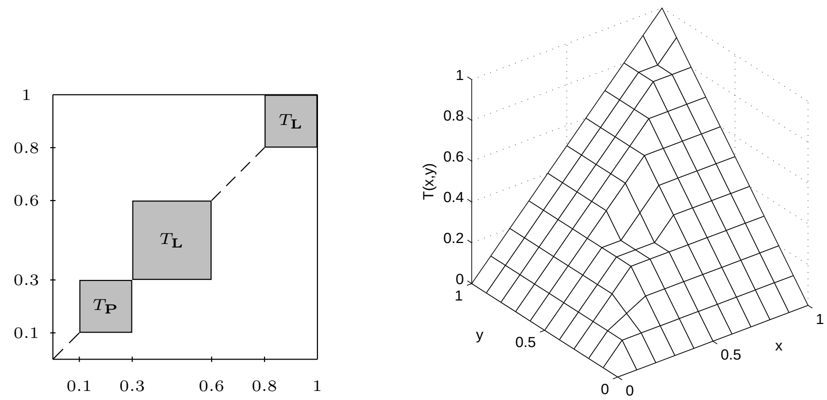


Figure 2.1: Construction of ordinal sums: the important parts of the domain of $T = (\langle 0.1, 0.3, T_{\mathbf{P}} \rangle, \langle 0.3, 0.6, T_{\mathbf{L}} \rangle, \langle 0.8, 1, T_{\mathbf{L}} \rangle)$ are marked on the left and its $3D$ plot on the right.

Definition 2.7 *A fuzzy relation R is called T-transitive (with T an arbitrary t-norm) if for any subsets A, B and C of X the following inequality holds:*

$$T(R(A,B), R(B,C)) \leq R(A,C)\,.$$

Note that T-transitivity implies T'-transitivity for all $T' \leq T$.

The $T_{\mathbf{M}}$-transitivity (min-transitivity) property reads as

$$\min(R(A,B), R(B,C)) \leq R(A,C)\,. \tag{2.6}$$

Note that a similarity measure which is $T_{\mathbf{M}}$-transitive is also called a *similarity relation*. An important theorem states that R is a similarity relation if and only if every α-cut R_α is a crisp equivalence relation (i.e. a reflexive, symmetric and min-transitive binary relation) [96].

The $T_{\mathbf{M}}$-transitivity property is well known in the field of hierarchical clustering as it plays an essential role in the construction of a partition tree. If R is a similarity relation, then for each α, R_α is an equivalence relation on $X = \{x_1, \ldots, x_n\}$. The equivalence classes of x_i with respect to R_α constitute the partition of X at cutting level α. It is clear that with decreasing α, the equivalence classes tend to merge. The graph representation of this hierarchy of equivalence classes is called the partition tree.

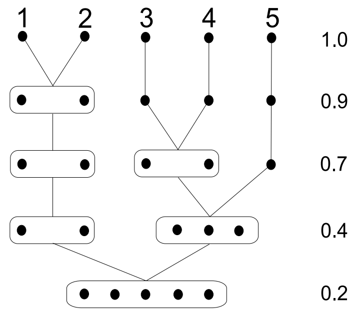


Figure 2.2: The partition tree associated to the $T_{\mathbf{M}}$-transitive matrix A_S.

Example 2.4 Consider the similarity relation S on $X = \{1, 2, 3, 4, 5\}$ with matrix representation

$$A_S = \begin{bmatrix} 1 & 0.9 & 0.2 & 0.2 & 0.2 \\ 0.9 & 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.7 & 0.4 \\ 0.2 & 0.2 & 0.7 & 1 & 0.4 \\ 0.2 & 0.2 & 0.4 & 0.4 & 1 \end{bmatrix}.$$

The partition tree associated to A_S is depicted in Figure 2.2.

More specifically, for a $T_{\mathbf{M}}$-transitive similarity measure S the mapping $d = 1 - S$ is an ultrametric. Unfortunately, the family of similarity measures (2.4) does not contain any $T_{\mathbf{M}}$-transitive members.

However, also $T_{\mathbf{L}}$- and $T_{\mathbf{P}}$-transitive similarity measures are of interest due to their correspondence with $[0, 1]$-valued pseudometrics [23, 24]:

(i) A similarity measure S is $T_{\mathbf{L}}$-transitive (Łukasiewicz-transitive) if it holds that

$$S(A, B) + S(B, C) - 1 \leq S(A, C)\,. \tag{2.7}$$

It then holds that the mapping $d = 1 - S$ is a pseudo-metric, i.e. the triangle inequality holds:

$$d(A, C) \leq d(A, B) + d(B, C)\,.$$

(ii) A similarity measure S is $T_{\mathbf{P}}$-transitive (product-transitive, a property stronger than $T_{\mathbf{L}}$-transitivity) if it holds that

$$S(A, B) \cdot S(B, C) \leq S(A, C)\,. \tag{2.8}$$

It now holds that the mapping $d = -\log S$ is a pseudo-metric.

The transitivity of the similarity measures in Table 2.6 and the inclusion measures in Table 2.7 is indicated in the last column. The $T_{\mathbf{L}}$-transitive, $T_{\mathbf{P}}$-transitive and $T_{\mathbf{M}}$-transitive members of the families of similarity measures (2.4) and inclusion measures (2.5) are identified in the following sections.

2.6 T-transitive similarity measures

2.6.1 Łukasiewicz-transitive members

In this subsection, we characterize the Łukasiewicz-transitive members of family (2.4).

Theorem 2.1 *The $T_{\mathbf{L}}$-transitive members of family (2.4) are characterized by the necessary and sufficient condition*

$$x' \geq \max(y,\, z)\,. \tag{2.9}$$

Proof. In order to identify the conditions on the parameters x, x', y and z, we have to verify when inequality (2.7) is fulfilled. Consider the setting in Figure 2.3, then the following equalities hold:

$$\begin{aligned}
|A \setminus B| &= a_1 + b_2\,, & |A \cap B| &= b_3 + c\,,\\
|A \setminus C| &= a_1 + b_3\,, & |A \cap C| &= b_2 + c\,,\\
|B \setminus C| &= a_2 + b_3\,, & |B \cap C| &= b_1 + c\,,\\
|A \triangle B| &= a_1 + a_2 + b_1 + b_2\,, & |(A \cup B)^c| &= a_3 + d\,,\\
|A \triangle C| &= a_1 + a_3 + b_1 + b_3\,, & |(A \cup C)^c| &= a_2 + d\,,\\
|B \triangle C| &= a_2 + a_3 + b_2 + b_3\,, & |(B \cup C)^c| &= a_1 + d\,,
\end{aligned}$$

and inequality (2.7) can be rewritten as

$$\begin{aligned}
(x'-x) \quad &\Bigg(-\frac{a_1+a_3+b_1+b_3}{x'(a_1+a_3+b_1+b_3)+y(b_2+c)+z(a_2+d)}\\
&+\frac{a_1+a_2+b_1+b_2}{x'(a_1+a_2+b_1+b_2)+y(b_3+c)+z(a_3+d)}\\
&+\frac{a_2+a_3+b_2+b_3}{x'(a_2+a_3+b_2+b_3)+y(b_1+c)+z(a_1+d)}\Bigg) \geq 0\,.
\end{aligned} \tag{2.10}$$

Since $x' > x$, we can omit the factor $x' - x$.

Setting $a_2 = b_2 = c = d = 0$, we obtain the following inequality

$$-\frac{1}{x'} + \frac{a_1+b_1}{x'(a_1+b_1)+yb_3+za_3} + \frac{a_3+b_3}{x'(a_3+b_3)+yb_1+za_1} \geq 0\,.$$

Converting the fractions of the left-hand side of this inequality such that they have a common (positive) denominator, it suffices to require

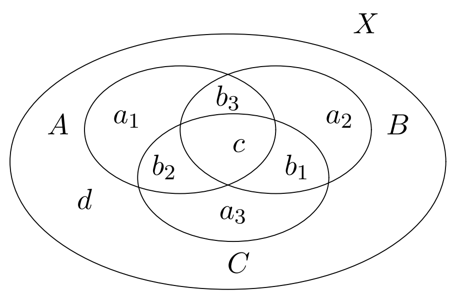


Figure 2.3: Notations for cardinalities associated with three ordinary subsets A, B and C of a finite universe X.

the positivity of the numerator, leading to the condition

$$(x'^2 - z^2)a_1a_3 + (x'^2 - y^2)b_1b_3 + (x'^2 - yz)(a_1b_3 + a_3b_1) \geq 0\,.$$

In particular, setting $a_1 = a_3 = 0$, we obtain

$$(x'^2 - y^2)b_1b_3 \geq 0\,.$$

This inequality is only fulfilled when $x' \geq y$. In the same way, we can set $b_1 = b_3 = 0$, leading to the condition $x' \geq z$. Other combinations do not lead to further conditions on x', y and z.

Similarly, we can start by setting $a_1 = b_1 = c = d = 0$ or $a_3 = b_3 = c = d = 0$, but none of these choices leads to new conditions. We conclude that $x' \geq \max(y, z)$ is a necessary condition for inequality (2.7) to hold.

If we carefully expand inequality (2.10), it is easy to see that $x' \geq \max(y, z)$ is also a sufficient condition, since in that case no negative terms occur in the expanded expression. □

Corollary 2.1 *The $T_{\mathbf{L}}$-transitive members of the ratio model (2.2) are characterized by $\alpha = \beta \geq 1$.*

Corollary 2.2 *The $T_{\mathbf{L}}$-transitive members of the family S_θ are characterized by $\theta \geq 1$ and those of the family T_ϕ by $\phi \geq 1$.*

Remark that Gower and Legendre already proved that $\theta \geq 1$ (resp. $\phi \geq 1$) is a sufficient condition for $1 - S_\theta$ (resp. $1 - T_\phi$) to satisfy the triangle inequality, but left it open whether this condition is also necessary. This is now clear.

2.6.2 Product-transitive members

In this subsection, we identify the product-transitive members of family (2.4).

Theorem 2.2 *The $T_{\mathbf{P}}$-transitive members of family (2.4) are characterized by the necessary and sufficient condition*

$$x\,x' \geq \max(y^2, z^2)\,. \tag{2.11}$$

Proof. In order to identify the conditions on the parameters x, x', y and z, we have to verify when inequality (2.8) is fulfilled. Considering the setting in Figure 2.3, the latter inequality reads

$$\begin{aligned}(x'-x)\quad &\left(-\frac{a_1+a_3+b_1+b_3}{N_1}+\frac{a_1+a_2+b_1+b_2}{N_2}+\frac{a_2+a_3+b_2+b_3}{N_3}\right.\\ &\left.-\frac{(x'-x)(a_1+a_2+b_1+b_2)(a_2+a_3+b_2+b_3)}{N_2N_3}\right)\geq 0\,,\end{aligned} \tag{2.12}$$

with

$$\begin{aligned}N_1 &= x'(a_1+a_3+b_1+b_3)+y(b_2+c)+z(a_2+d)\,,\\ N_2 &= x'(a_1+a_2+b_1+b_2)+y(b_3+c)+z(a_3+d)\,,\\ N_3 &= x'(a_2+a_3+b_2+b_3)+y(b_1+c)+z(a_1+d)\,.\end{aligned}$$

Again, we can omit the factor $x'-x$. Setting $a_2=b_2=c=d=0$, we obtain the following inequality

$$\begin{aligned}&-\frac{1}{x'}+\frac{a_1+b_1}{x'(a_1+b_1)+yb_3+za_3}+\frac{a_3+b_3}{x'(a_3+b_3)+yb_1+za_1}\\ &-\frac{(x'-x)(a_1+b_1)(a_3+b_3)}{(x'(a_1+b_1)+yb_3+za_3)(x'(a_3+b_3)+yb_1+za_1)}\geq 0\,.\end{aligned}$$

Converting the fractions of the left-hand side of this inequality such that they have a common (positive) denominator, it suffices to require the positivity of the numerator, leading to the condition

$$(xx'-z^2)a_1a_3+(xx'-y^2)b_1b_3+(xx'-yz)(a_1b_3+a_3b_1)\geq 0\,.$$

In particular, setting $a_1=a_3=0$, we obtain

$$b_1b_3(x\,x'-y^2)\geq 0\,.$$

This inequality is only fulfilled when $x\,x' \geq y^2$. In the same way, we can set $b_1 = b_3 = 0$, leading to the condition $x\,x' \geq z^2$. Setting $a_1 = b_3 = 0$ or $a_3 = b_1 = 0$ leads to the condition $x\,x' \geq y\,z$, which is weaker than the two previous conditions. Other combinations do not lead to further conditions on x, x', y and z. We conclude that $x\,x' \geq \max(y^2, z^2)$ is a necessary condition for inequality (2.8) to hold.

Again, if we carefully expand inequality (2.12), it is easy to see that $x\,x' \geq \max(y^2, z^2)$ is also a sufficient condition. □

Corollary 2.3 *The ratio model (2.2) does not contain $T_{\mathbf{P}}$-transitive members.*

Corollary 2.4 *The families S_θ and T_ϕ do not contain $T_{\mathbf{P}}$-transitive members.*

2.6.3 Min-transitive members

In this subsection, we prove that the family of cardinality-based similarity measures (2.4) does not contain any min-transitive members.

Theorem 2.3 *Family (2.4) does not contain any $T_{\mathbf{M}}$-transitive members.*

Proof. We have to verify when inequality (2.6) is fulfilled. First, we verify when the following inequality is fulfilled: $S(A,B) \leq S(A,C)$.

Considering the setting as in Figure 2.3, the latter inequality reads

$$\frac{a_1 + a_2 + b_1 + b_2}{N_1} - \frac{a_1 + a_3 + b_1 + b_3}{N_2} \geq 0\,, \tag{2.13}$$

with

$$\begin{aligned} N_1 &= x'(a_1 + a_2 + b_1 + b_2) + y(b_3 + c) + z(a_3 + d)\,, \\ N_2 &= x'(a_1 + a_3 + b_1 + b_3) + y(b_2 + c) + z(a_2 + d)\,. \end{aligned}$$

Since inequality (2.13) should hold for any a_i, b_i, c and d, with $i \in \{1,2,3\}$, it should hold for $a_2 = b_2 = 0$, $c = d = 0$, $a_i \neq 0$ and $b_i \neq 0$ for $i \in \{1,3\}$ in particular. In that case, inequality (2.13) can be rewritten as:

$$\frac{a_1 + b_1}{x'(a_1 + b_1) + yb_3 + za_3} - \frac{1}{x'} \geq 0\,.$$

Converting the fractions of the left-hand side of this inequality such that they have a common (positive) denominator, it suffices to require the positivity of the numerator, leading to the condition

$$-yb_3 - za_3 \geq 0\,,$$

which is obviously a contradiction. □

2.7 T-transitive inclusion measures

The rational inclusion measures introduced in [21] have also been investigated for properties such as monotonicity and T-transitivity. An overview of some cardinality-based inclusion measures and their transitivity properties can be found in Table 2.7.

2.7.1 Łukasiewicz-transitive members

In this subsection, we characterize the Łukasiewicz-transitive members of family (2.5).

Theorem 2.4 *The $T_{\mathbf{L}}$-transitive members of the class of inclusion measures (2.5) are characterized by*

$$x' \geq \max(y, z)\,. \tag{2.14}$$

Proof. In order to identify the conditions on the parameters x, y, z and x' in (2.5), we have to verify when inequality (2.7) is fulfilled. Considering the setting as in Figure 2.3, then the following equalities hold:

$$\begin{aligned}
|A \setminus B| &= a_1 + b_2\,, & |A \cap B| &= b_3 + c\,,\\
|A \setminus C| &= a_1 + b_3\,, & |A \cap C| &= b_2 + c\,,\\
|B \setminus C| &= a_2 + b_3\,, & |B \cap C| &= b_1 + c\,,\\
|A \triangle B| &= a_1 + a_2 + b_1 + b_2\,, & |(A \cup B)^c| &= a_3 + d\,,\\
|A \triangle C| &= a_1 + a_3 + b_1 + b_3\,, & |(A \cup C)^c| &= a_2 + d\,,\\
|B \triangle C| &= a_2 + a_3 + b_2 + b_3\,, & |(B \cup C)^c| &= a_1 + d\,,
\end{aligned}$$

and inequality (2.7) can be rewritten as

$$\begin{aligned}
(x' - x) \quad & \left(- \frac{a_1 + b_3}{x'(a_1 + a_3 + b_1 + b_3) + y(b_2 + c) + z(a_2 + d)} \right. \\
& + \frac{a_1 + b_2}{x'(a_1 + a_2 + b_1 + b_2) + y(b_3 + c) + z(a_3 + d)} \\
& \left. + \frac{a_2 + b_3}{x'(a_2 + a_3 + b_2 + b_3) + y(b_1 + c) + z(a_1 + d)} \right) \geq 0\,.
\end{aligned} \tag{2.15}$$

Since $x' > x$, we can omit the factor $x' - x$. In particular, putting $a_2 = b_2 = c = d = 0$, we obtain the following inequality

$$\begin{aligned}
&- \frac{a_1 + b_3}{x'(a_1 + a_3 + b_1 + b_3)} + \frac{a_1}{x'(a_1 + b_1) + yb_3 + za_3} \\
&+ \frac{b_3}{x'(a_3 + b_3) + yb_1 + za_1} \geq 0\,.
\end{aligned}$$

If we reduce the left-hand side of this inequality to the same denominator, it is sufficient to study the numerator, since the denominator is always positive. In particular, if we put $a_1 = a_3 = 0$, then the latter inequality reduces to

$$b_1 b_3 (x' b_1 + y b_3)(x' - y) \geq 0 \,.$$

This inequality is only fulfilled if $x' \geq y$. In the same way, we can put $b_1 = b_3 = 0$, which leads to the condition $x' \geq z$. Other combinations do not lead to different conditions on x', y and z. In a similar way, we can also put $a_1 = b_1 = c = d = 0$ or $a_3 = b_3 = c = d = 0$, but none of them leads to new conditions on x', y and z. We conclude that $x' \geq \max(y, z)$ is a necessary condition for inequality (2.7) to hold.

If we expand inequality (2.15), it is easy to see that $x' \geq \max(y, z)$ is also a sufficient condition, since no negative terms occur in the expanded expression. □

2.7.2 Product-transitive members

Next, we derive the necessary conditions such that the members of the family of inclusion measures (2.5) are $T_{\mathbf{P}}$-transitive. Note that we are not able to identify a set of conditions that are at the same time necessary and sufficient such that the members of family (2.5) are $T_{\mathbf{P}}$-transitive.

Theorem 2.5 *Necessary conditions on x, y, z and x' in order that the members of the class of inclusion measures (2.5) are $T_{\mathbf{P}}$-transitive, are given by*

$$x' \geq \max(y, z) \;\wedge\; x\,x' \geq \max(y\,z, z(x' - z), y(x' - y)) \,. \tag{2.16}$$

Sufficient conditions on x, y, z and x' in order that the members of the class of inclusion measures (2.5) are $T_{\mathbf{P}}$-transitive, are given by

$$\begin{aligned} & x' \geq \max(y, z) \\ \wedge\;\; & x\,x' \geq \max(y\,z, z(x' - z), y(x' - y)) \\ \wedge\;\; & y^2 - x'^2 + x\,x' + x'\,z + x\,z + x'\,y \geq 0 \\ \wedge\;\; & z^2 - x'^2 + x\,x' + x'\,z + x\,y + x'\,y \geq 0 \,. \end{aligned} \tag{2.17}$$

Proof. In order to identify the conditions on the parameters x, y, z and x' in (2.5), we have to verify when inequality (2.8) is fulfilled. Again, considering the setting as in Figure 2.3, then the latter inequality reduces to

$$(x'-x)\ \left(-\frac{a_1+b_3}{N_1}+\frac{a_1+b_2}{N_2}+\frac{a_2+b_3}{N_3}\right.\\ \left.-\frac{(x'-x)(a_1+b_2)(a_2+b_3)}{N_2N_3}\right)\geq 0\,, \tag{2.18}$$

with

$$\begin{aligned}N_1 &= x'(a_1+a_3+b_1+b_3)+y(b_2+c)+z(a_2+d)\,,\\ N_2 &= x'(a_1+a_2+b_1+b_2)+y(b_3+c)+z(a_3+d)\,,\\ N_3 &= x'(a_2+a_3+b_2+b_3)+y(b_1+c)+z(a_1+d)\,.\end{aligned}$$

Again, we omit the factor $x'-x$, since $x'>x$. In particular, putting $a_2=b_2=c=d=0$, we obtain the following inequality

$$\begin{aligned}&-\ \frac{a_1+b_3}{x'(a_1+a_3+b_1+b_3)}+\frac{a_1}{x'(a_1+b_1)+yb_3+za_3}\\ &+\ \frac{b_3}{x'(a_3+b_3)+yb_1+za_1}\\ &-\ \frac{(x'-x)a_1b_3}{(x'(a_1+b_1)+yb_3+za_3)(x'(a_3+b_3)+yb_1+za_1)}\geq 0\,.\end{aligned}$$

If we convert the fractions of the left-hand side of this inequality so that they have a common denominator, it is sufficient to study the numerator, since the denominator is always positive. In particular, if we put $b_1=a_3=0$, then the latter inequality reduces to

$$a_1b_3(xx'-yz)\geq 0\,.$$

This inequality is only fulfilled if $xx'\geq yz$. Putting $a_3=b_3=c=d=0$ and $b_1=b_2=0$, we obtain the following inequality

$$a_2((z^2-x'z+xx')a_1^2+z(x'+x)a_1a_2+x'za_2^2)\geq 0\,.$$

This inequality is only fulfilled if $xx'\geq x'z-z^2$. This can be seen by considering a constant $a_2>0$, and by letting $n\to\infty$ and $a_1\to\infty$, as the result should hold for any n. In a similar way, we obtain $xx'\geq x'y-y^2$. Other combinations do not lead to different conditions on x,x',y and z. We conclude that conditions (2.16) are necessary conditions for inequality (2.8) to hold. Note that they are not sufficient for $T_{\mathbf{P}}$-transitivity.

If we expand inequality (2.18), then it is easy to see that

$$\begin{aligned} & x' \geq \max(y, z) \\ \wedge \quad & x\,x' \geq \max(y\,z, z(x'-z), y(x'-y)) \\ \wedge \quad & x'^2 - y^2 + x\,x' - x'\,z \geq 0 \\ \wedge \quad & x'^2 - z^2 + x\,x' - x'\,y \geq 0 \\ \wedge \quad & y^2 - x'^2 + x\,x' + x'\,z + x\,z + x'\,y \geq 0 \\ \wedge \quad & z^2 - x'^2 + x\,x' + x'\,z + x\,y + x'\,y \geq 0 \\ \wedge \quad & x + x' \geq y + z \end{aligned}$$

are sufficient conditions. Note that conditions $x\,x' \geq y\,z$ and $x' \geq \max(y, z)$ imply $x'^2 - y^2 + x\,x' - x'\,z \geq 0$, $x'^2 - z^2 + x\,x' - x'\,y \geq 0$ and $x + x' \geq y + z$. Therefore, the latter conditions can be omitted from the list of sufficient conditions (2.17). $\square$

Note that it is possible to find values for the parameters x, x', y and z such that the set of necessary conditions (2.16) are fulfilled, but not the set of sufficient conditions (2.17) and that the corresponding inclusion measure is $T_{\mathbf{P}}$-transitive. For example, putting $x = 5$, $x' = 10$, $y = 1$ and $z = 0$ leads to the following inclusion measure:

$$I(A,B) = \frac{5|A \setminus B| + 10|B \setminus A| + |A \cap B|}{10|A \triangle B| + |A \cap B|}\,. \tag{2.19}$$

It is easy to verify that the left part of the following inequality

$$z^2 - x'^2 + x\,x' + x'\,z + x\,y + x'\,y \geq 0 \tag{2.20}$$

equals -40 and therefore, inequality (2.20) is obviously not fulfilled. Analogously, the left part of the following inequality

$$y^2 - x'^2 + x\,x' + x'\,z + x'\,y + x\,z \geq 0 \tag{2.21}$$

equals -39 and therefore, inequality (2.20) is also not fulfilled. Next we prove that the inclusion measure (2.19) is $T_{\mathbf{P}}$-transitive.

Property 2.1 *Inclusion measure (2.19) is $T_{\mathbf{P}}$-transitive.*

Proof. We have to verify when inequality (2.8) is fulfilled. Consider the setting as in Figure 2.3, then the latter inequality reduces to

$$1 - \frac{5(a_1+b_3)}{N_1} - \left(1 - \frac{5(a_1+b_2)}{N_2}\right)\left(1 - \frac{5(a_2+b_3)}{N_3}\right) \geq 0\,, \tag{2.22}$$

with

$$\begin{aligned} N_1 &= 10a_1 + 10a_3 + 10b_1 + 10b_3 + b_2 + c\,, \\ N_2 &= 10a_1 + 10a_2 + 10b_1 + 10b_2 + b_3 + c\,, \\ N_3 &= 10a_2 + 10a_3 + 10b_2 + 10b_3 + b_1 + c\,. \end{aligned}$$

If we convert the fractions of the left-hand side of this inequality so that they have a common denominator, it is sufficient to study the numerator, since the denominator is always positive. The numerator is a function of 7 variables and will be denoted by $f(a_1, a_2, a_3, b_1, b_2, b_3, c)$. Since all terms including b_1, a_3 or c are positive, the following holds

$$\begin{aligned} f(a_1, a_2, a_3, b_1, b_2, b_3, c) &\geq f(a_1, a_2, 0, 0, b_2, b_3, 0) \\ &= 5\,[\,41b_2b_3^2 + 10a_2^2b_2 - 39a_2b_2b_3 \\ &- 35a_1a_2b_2 + 50a_1^2a_2 + 15a_2b_2^2 \\ &+ 50a_1b_3^2 + 50a_1a_2b_3 + 15b_2^2b_3 + 50a_1^2b_3 \\ &+ 55a_1b_2b_3 + 10b_2^3 + 10a_1b_2^2\,] \\ &= 5\,[\,b_2(\sqrt{41}b_3 - \sqrt{10}a_2)^2 + (2\sqrt{410} - 39)a_2b_2 \\ &+ 5a_2(\sqrt{10}a_1 - \sqrt{3}b_2)^2 + (2\sqrt{30} - 7)a_1b_2 \\ &+ 50a_1b_3^2 + 50a_1a_2b_3 + 15b_2^2b_3 + 50a_1^2b_3 \\ &+ 55a_1b_2b_3 + 10b_2^3 + 10a_1b_2^2\,]\,. \end{aligned}$$

Since $2\sqrt{410} - 39 \geq 0$ and $2\sqrt{30} - 7 \geq 0$, $f(a_1, a_2, a_3, b_1, b_2, b_3, c) \geq 0$. This completes our proof. □

2.7.3 Min-transitive members

In this subsection, we prove that the family of cardinality-based inclusion measures (2.5) does not contain any min-transitive members.

Theorem 2.6 *Family (2.5) does not contain any $T_{\mathbf{M}}$-transitive members.*

Proof. We have to verify when inequality (2.6) is fulfilled. First, we verify when the following inequality is fulfilled: $I(A, B) \leq I(A, C)$.

Consider the setting as in Figure 2.3, the latter inequality reads

$$(x' - x)\left(\frac{a_1 + b_2}{N_1} - \frac{a_1 + b_3}{N_2}\right) \geq 0\,, \tag{2.23}$$

with

$$N_1 = x'(a_1 + a_2 + b_1 + b_2) + y(b_3 + c) + z(a_3 + d)\,,$$
$$N_2 = x'(a_1 + a_3 + b_1 + b_3) + y(b_2 + c) + z(a_2 + d)\,.$$

Again, we omit the factor $x' - x$, since $x' > x$. Since inequality (2.23) should hold for any a_i, b_i, c and d, with $i \in \{1,2,3\}$, it should hold for $a_1 = a_2 = b_2 = 0$, $c = d = 0$, $a_3 \neq 0$ and $b_i \neq 0$ for $i \in \{1,3\}$ in particular. In that case, inequality (2.23) can be rewritten as:

$$-\frac{b_3}{x'(a_3 + b_1 + b_3)} \geq 0\,,$$

which is obviously a contradiction. □

2.8 Conclusions and indications for future research

In this chapter, we have given an overview of the comparison measures for binary vectors frequently used in practice. As binary vectors can be identified with ordinary sets, it is natural to convert these comparison measures for binary vectors into measures for ordinary sets. The latter are then based on the cardinalities of the sets involved. We mentioned three already existing parametric families of cardinality-based similarity measures, one family of cardinality-based inclusion measures and introduced two new parametric families. Moreover, we have identified the $T_{\mathbf{L}}$-transitive and $T_{\mathbf{P}}$-transitive members of these families. Finally, we concluded that none of these families have $T_{\mathbf{M}}$-transitive members.

In this thesis, we only consider a family of similarity measures based on the addition (or substraction) of cardinalities. However, also similarity measures which rely on the multiplication of cardinalities (for example, the Cosine coefficient) exist. It is clear that these measures cannot be a member of the suggested family of similarity measures and that introducing a new family, which covers the latter type of similarity measures, is within the bounds of possibilities.

The parameters x, y, z and x' of the family of similarity measures are a second limitation. The use of only four parameters causes that some similarity measures (for instance, the first coefficient of Kulczynski) are put out of action. This problem could be solved by considering six parameters instead of four. However, this new direction also needs further research.

Chapter 3

Fuzzification schemes

3.1 Introduction

In Chapters 6 and 7, we will fuzzify the parametrized families of cardinality-based similarity and inclusion measures which were defined in the previous chapter. Since these parametrized families are based on the cardinality of an ordinary set and the basic classical set operations, such as intersection, union and (symmetric) difference, we need to formulate fuzzification schemes in order to translate these operations to their fuzzy counterpart.

This chapter is organized as follows. The basic scalar cardinality of a fuzzy set is introduced in Section 3.2 as well as different definitions of fuzzy cardinalities. Several models to fuzzify the intersection, union and (symmetric) difference of two ordinary sets are explained in Section 3.3 of this chapter.

Before we proceed, we introduce the notions of a normal fuzzy set, a convex fuzzy set and an α-cut of a fuzzy set. Consider a finite universe $X = \{x_1, \ldots, x_n\}$. A fuzzy set A in X is said to be normal if and only if there exists an $x_i \in X$, for any $i = 1, \ldots, n$, such that $A(x_i) = 1$. A fuzzy set A in X is convex if and only if for any $x_1, x_2 \in X$ the following holds: $A(\lambda x_1 + (1 - \lambda)x_2) \geq \min(A(x_1), A(x_2))$, for any λ belonging to the unit interval. An α-cut of a fuzzy set A is a crisp set A_α, defined by $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$.

3.2 Cardinality of a fuzzy set

Measuring the cardinality of a fuzzy set is a necessary task in many problems, such as fuzzy querying in databases, expert systems, evaluation of natural language statements, and many others. With an eye to the fuzzification of parametric families (2.4) and (2.5) of cardinality-based similarity and inclusion measures, we also need a concept of the cardinality of a fuzzy set.

If we want to count elements in a fuzzy set, the main problem is that those elements only belong to a fuzzy set to a certain degree. Nevertheless, it would be useful to have at one's disposal a counterpart for the cardinality of an ordinary set. As is often the case when extending a classical set-theoretical notion to fuzzy set theory, various alternative definitions for the notion of the cardinality of a fuzzy set (on a finite universe) have been proposed. These proposals can be roughly subdivided in three categories:

(i) *scalar cardinalities*, where the cardinality of a fuzzy set is a positive real number (Casasnovas and Torrens [14], De Luca and Termini [25], Gottwald [38], Kaufmann [49], Wygralak [91]),

(ii) *fuzzy cardinalities*, where the cardinality of a fuzzy set is defined as a fuzzy quantity (not necessarily convex) (Blanchard [5], Delgado *et al.* [27], Dubois and Prade [30], Wygralak [91], Zadeh [95]), and

(iii) *integer cardinalities*, obtained by defuzzifying in some sense fuzzy cardinalities (Ralescu [68]) (which essentially reduces to cardinalities of 0.5-cuts).

In the first approach, the cardinality of a fuzzy set A, denoted by $|A|$, is a positive real number. Therefore, this type of cardinality is called a scalar cardinality. The most basic definition of the scalar cardinality of a fuzzy set A on a finite universe $X = \{x_1, \ldots, x_n\}$ proposed by De Luca and Termini [25] is the following:

$$|A| = \sum_{i=1}^{n} A(x_i)\,.$$

This scalar cardinality is often referred to as the sigma count or sometimes the power of a fuzzy set. We will simply refer to it as the *basic scalar cardinality*. It goes without saying that this way of defining the cardinality of a fuzzy set is rather simplistic, but on the other hand very convenient to use.

The second approach offers a fuzzy perception of cardinality: it is again a fuzzy set. In the following, we give an overview of the definitions available in the literature, which can also be found in [27]. A first definition of fuzzy cardinality is due to Zadeh [97]:

$$|A|(i) = \sup\{\alpha \mid |A_\alpha| = i\}\,, \quad i = 0, \ldots, n\,, \tag{3.1}$$

The main problem that has been attributed to this method is that the valuation property (i.e. $|A \cup B| + |A \cap B| = |A| + |B|$) is not fulfilled as $|A|$ is not a convex fuzzy set [30]. To recover the valuation property, Dubois and Prade suggested the following definition of the cardinality of a fuzzy set A [30]:

$$|A|(i) = \sup\{\alpha \mid |A_\alpha| \geq i\}\,, \quad i = 0, \ldots, n\,.$$

This definition is in fact nothing more than $\mathrm{FGCount}(A)$, defined by Zadeh [98]. The main drawback of this definition is that $\mathrm{FGCount}$ does not match exactly the idea of cardinality, since for crisp sets, $\mathrm{FGCount}$ provides a set of integers $\{0, 1, \ldots, |A|\}$. From a practical point of view, this is not satisfactory. To avoid this problem, Zadeh proposed $\mathrm{FECount}(A)$ as the fuzzy cardinality of A, defined as:

$$\mathrm{FECount}(A) = \mathrm{FGCount}(A) \cap \mathrm{FLCount}(A)\,,$$

where

$$\mathrm{FLCount}(A)(i) = \sup\{\alpha \mid |A_\alpha| \geq n - i\}\,, \quad i = 0, \ldots, n\,.$$

An equivalent expression for $\mathrm{FECount}(A)$ was introduced by Wygralak [91] and also Ralescu has used the same definition in [68]. $\mathrm{FECount}(A)$ is a convex fuzzy set, but not normal. Wygralak [91, 92] showed that the valuation property is fulfilled if and only if the intersection and union of two fuzzy sets are modelled by min and max, respectively.

Dubois and Prade proposed to define the cardinality of a fuzzy set as follows,

$$\|A\| = |A| \cap \{|A_1|, |A_1| + 1, |A_1| + 2, \ldots\}\,,$$

with $|A|$ defined as in (3.1). The fuzzy cardinality $\|A\|$ is a normalized, convex fuzzy set and moreover, the valuation property is fulfilled if and only if the intersection and union of two fuzzy sets are modelled by the operators min and max, respectively.

In the third approach, the cardinality of a fuzzy set A is obtained by defuzzifying in some sense fuzzy cardinalities. Ralescu [68] defines the cardinality of a fuzzy set A as follows:

$$|A| = \begin{cases} 0 & \text{if } A = \emptyset \\ j & \text{if } A \neq \emptyset \text{ and } A(j) \geq 0.5 \\ j-1 & \text{if } A \neq \emptyset \text{ and } A(j) < 0.5 \end{cases}$$

where $j = \max\{1 \geq s \geq n \mid A(s-1) + A(s) \geq 1\}$. But, this definition is nothing more than counting the elements of the α-cut of A, with $\alpha = 0.5$ and therefore $|A| = |A_{0.5}|$.

From what precedes, we conclude that there are many different ways of defining the cardinality of a fuzzy set, but we restrict ourselves through the remainder of this work, to the use of the basic scalar cardinality only.

3.3 Translation of classical set operations

We do not want to restrict ourselves to the study of a parametric family of cardinality-based measures for ordinary sets only, but also want to provide a family of fuzzy similarity (or inclusion) measures. Therefore, we need to fuzzify the parametric families introduced in Chapter 2. For this fuzzification process, we have two possibilities:

(i) rewrite the families for ordinary sets in terms of intersections only and fuzzify this new expression, or

(ii) establish fuzzification rules for the complement of a fuzzy set, fuzzy set union and fuzzy set (symmetrical) difference.

In both cases, we use the basic scalar cardinality for the cardinality of a fuzzy set and also need to fuzzify the intersection of two fuzzy sets.

3.3.1 Intersection of two fuzzy sets

As usual, we define the intersection of two fuzzy sets A and B on a finite universe X pointwisely, i.e. $A \cap B(x) = I(A(x), B(x))$, by means of an appropriate function I that generalizes Boolean conjunction. Since we will intersect at most two fuzzy sets at the same time, it suffices to consider as suitable I a commutative conjunctor.

Definition 3.1 *A binary operation $I : [0,1]^2 \to [0,1]$ is called a conjunctor if it satisfies:*

(i) *Neutral element* 1*:* $(\forall x \in [0,1])(I(x,1) = I(1,x) = x)$.

(ii) *Monotonicity: I is increasing in each variable.*

Note that any conjunctor I coincides on $\{0,1\}^2$ with the Boolean conjunction and satisfies:

(i′) Absorbing element 0: $(\forall x \in [0,1])(I(x,0) = I(0,x) = 0)$.

Moreover, any conjunctor I is bounded from above by the minimum operator $T_{\mathbf{M}}$, i.e. $I(x,y) \leq T_{\mathbf{M}}(x,y) = \min(x,y)$.

In this thesis, we focus our attention on three particular classes of conjunctors: the class of triangular norms (t-norms), the class of copulas and the class of quasi-copulas. Where t-norms, introduced in Section 2.5, have the additional properties of associativity and commutativity [52], copulas have the property of moderate growth, while quasi-copulas have the 1-Lipschitz property.

Copulas were introduced by Sklar in 1959 [77] and are used for combining marginal probability distributions into joint probability distributions. In short, Sklar showed that if H is a bivariate distribution function with margins $F(x)$ and $G(y)$, then there exists a copula C such that $H(x,y) = C(F(x), G(y))$. They are not only of interest in the field of probability and statistics (as a way of studying measures of dependence or for constructing families of bivariate distributions), but also for the "fuzzy community", copulas are of great value. Since associative copulas are special continuous t-norms, they are applied in several domains where t-norms play a role.

The notion of quasi-copulas was recently introduced by Alsina et al. [1] in order to characterize operations on distribution functions that can or cannot be derived from operations on random variables. The following definition of a quasi-copula is not the original one as introduced by Alsina et al., but the one of Genest et al. [37].

Definition 3.2 [37, 61] *A binary operation $C : [0,1]^2 \to [0,1]$ is called a quasi-copula if it satisfies:*

(i) *Neutral element* 1.

(i′) *Absorbing element* 0.

(ii) *Monotonicity: C is increasing in each variable.*

(iii) *1-Lipschitz property: for any* $(x_1, x_2, y_1, y_2) \in [0,1]^4$ *it holds that:*

$$|C(x_1, y_1) - C(x_2, y_2)| \leq |x_1 - x_2| + |y_1 - y_2| \,.$$

If instead of (iii) C *satisfies*

(iv) *Moderate growth: for any* $(x_1, x_2, y_1, y_2) \in [0,1]^4$ *such that* $x_1 \leq x_2$ *and* $y_1 \leq y_2$ *it holds that:*

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

then C *is called a copula.*

Note that in case of a quasi-copula, condition (i') is superfluous, while for a copula condition (ii) can be omitted (as it follows from (iv) and (i')). As implied by the terminology used, any copula is a quasi-copula, and therefore has the 1-Lipschitz property; the opposite is, of course, not true. It is well known that a copula is a t-norm if and only if it is associative; conversely, a t-norm is a copula if and only if it is 1-Lipschitz. The three main continuous t-norms $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ are (associative and commutative) copulas.

In the remainder of this work, it is necessary to know that

Proposition 3.1 [37] *For any quasi-copula* C *it holds that* $T_{\mathbf{L}} \leq C \leq T_{\mathbf{M}}$.

We also need the concept of a stable quasi-copula. By analogy with the concept of a stable copula [53], we introduce the concept of the survival quasi-copula.

Definition 3.3 *The survival operator* $\hat{C}$ *associated to a quasi-copula* C *is the binary operation* $\hat{C}$ *defined by*

$$\hat{C}(x, y) = x + y - 1 + C(1 - x, 1 - y) \,.$$

We can prove that $\hat{C}$ is again a quasi-copula.

Proposition 3.2 *The survival operator* $\hat{C}$ *associated to a quasi-copula* C *is again a quasi-copula.*

Proof. It is easy to see that $\hat{C}$ has neutral element 1 and absorbing element 0. First, we prove that $\hat{C}$ is increasing. Suppose $x_1 \leq x_2$, then we should verify that the following inequality $\hat{C}(x_1, y) \leq \hat{C}(x_2, y)$ is fulfilled for any $y \in [0,1]$, or equivalently,

$$x_1 + C(1 - x_1, 1 - y) \leq x_2 + C(1 - x_2, 1 - y) \,. \tag{3.2}$$

Since C is a quasi-copula, C fulfills the 1-Lipschitz property:

$$\begin{aligned} & |C(1-x_1,1-y)-C(1-x_2,1-y)| \\ = & \ C(1-x_1,1-y)-C(1-x_2,1-y) \\ \leq & \ |1-x_1-1+x_2| = x_2-x_1 \, , \end{aligned}$$

therefore inequality (3.2) is fulfilled. In the same way, we can prove that $\hat{C}(x,y_1) \leq \hat{C}(x,y_2)$ for any $(x,y_1,y_2) \in [0,1]^3$, with $y_1 \leq y_2$.

Next, we show that $\hat{C}$ fulfills the 1-Lipschitz property.

$$\begin{aligned} & |\hat{C}(x_1,y_1)-\hat{C}(x_2,y_2)| \\ = & \ |\hat{C}(x_1,y_1)-\hat{C}(x_2,y_1)+\hat{C}(x_2,y_1)-\hat{C}(x_2,y_2)| \\ \leq & \ |\hat{C}(x_1,y_1)-\hat{C}(x_2,y_1)| + |\hat{C}(x_2,y_1)-\hat{C}(x_2,y_2)| \, . \end{aligned}$$

Suppose $x_1 \leq x_2$, then the following holds:

$$\begin{aligned} & |\hat{C}(x_1,y_1)-\hat{C}(x_2,y_1)| \\ = & \ \hat{C}(x_2,y_1)-\hat{C}(x_1,y_1) \\ = & \ x_2 + C(1-x_2,1-y_1) - x_1 - C(1-x_1,1-y_1) \\ \leq & \ x_2 - x_1 = |x_2-x_1| \end{aligned}$$

whereas the last inequality follows from $C(1-x_2,1-y_1)-C(1-x_1,1-y_1) \leq 0$.

In the same way, we can prove that $|\hat{C}(x_1,y_1)-\hat{C}(x_2,y_1)| \leq |x_2-x_1|$ holds when $x_1 > x_2$. And analogously,

$$|\hat{C}(x_2,y_1)-\hat{C}(x_2,y_2)| \leq |y_2-y_1| \, .$$

This completes our proof. □

Definition 3.4 *A quasi-copula C is called stable if $\hat{C} = C$.*

Example 3.1 The copula C defined by

$$C = \frac{T_{\mathbf{M}} + T_{\mathbf{L}}}{2}$$

is a stable copula [53].

One of the most important families of t-norms $\{T_\lambda^{\mathbf{F}}\}_{\lambda \in [0,\infty]}$ was obtained by Frank [35] as (continuous) solutions of the functional equation

$$T(x,y) + T^*(x,y) = x + y \, , \tag{3.3}$$

where T^* denotes the dual t-conorm of T defined by

$$T^*(x,y) = 1 - T(1-x, 1-y)\,.$$

A t-conorm can be defined as follows:

Definition 3.5 *[52] A binary operation $S : [0,1]^2 \to [0,1]$ is called a t-conorm if it satisfies:*

(i) *Neutral element* 0*:* $(\forall x \in [0,1])(S(x,0) = x)\,.$

(ii) *Monotonicity: S is increasing in each variable.*

(iii) *Commutativity:* $(\forall (x,y) \in [0,1]^2)(S(x,y) = S(y,x))\,.$

(iv) *Associativity:* $(\forall (x,y,z) \in [0,1]^3)(S(x,S(y,z)) = S(S(x,y),z))\,.$

The family $\{T^{\mathbf{F}}_\lambda\}_{\lambda\in[0,\infty]}$ of Frank t-norms (all of them are copulas) is given by

$$T^{\mathbf{F}}_\lambda(x,y) = \begin{cases} T_{\mathbf{M}}(x,y) & \text{if } \lambda = 0\,, \\ T_{\mathbf{P}}(x,y) & \text{if } \lambda = 1\,, \\ T_{\mathbf{L}}(x,y) & \text{if } \lambda = \infty\,, \\ \log_\lambda(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & \text{otherwise}\,. \end{cases}$$

The cases $\lambda \in \{0,1,\infty\}$ can be considered as limit cases of the general case. Note that any other t-norm fulfilling (3.3) is a particular kind of 'symmetric' ordinal sum of members of the Frank t-norm family, i.e. if $\langle a,b,T\rangle$ is a summand, then also $\langle 1-b, 1-a, T\rangle$ is a summand [52]. Moreover, in [53] the following proposition was proven concerning a stable, associative copula.

Proposition 3.3 *Let C be a copula. Then C is associative and stable (i.e. C fulfills Eq. (3.3)) if and only if there is a $\lambda \in [0,\infty]$ such that $C = T^{\mathbf{F}}_\lambda$ or if C is an ordinal sum of Frank t-norms of the form*

$$C = (\langle a_k, b_k, T^{\mathbf{F}}_{\lambda_k})_{k\in K}\rangle\,,$$

where for each $k \in K$ there is a $k' \in K$ such that $\lambda_k = \lambda_{k'}$ and $a_k + b_{k'} = a_{k'} + b_k = 1$.

3.3.2 Identities on fuzzy set cardinalities

Using the second option to fuzzify the parametric families introduced in Chapter 2, we need to establish fuzzification rules for the complement of a fuzzy set, fuzzy set union and fuzzy set (symmetric) difference. In this subsection we will propose a set of fuzzification rules such that identities on cardinalities of ordinary sets remain invariant when fuzzified.

A. The identity $|A^c| = n - |A|$

As we want to preserve the classical identity $|A^c| = n - |A|$, we should define the complement A^c of A as follows:

$$A^c(x_i) = 1 - A(x_i) \,.$$

B. The valuation property

In classical set theory, it holds that

$$|A \cap B| + |A \cup B| = |A| + |B| \,, \tag{3.4}$$

called the valuation property. If in fuzzy set theory Eq. (3.4) should hold, then $A \cup B$ should be defined by $A \cup B(x) = J(A(x), B(x))$, with J given by

$$J(x, y) = x + y - I(x, y) \,. \tag{3.5}$$

Proposition 3.4 *The operator J has the following properties:*

(i) J is a binary operation on $[0, 1]$ if and only if $I \geq T_{\mathbf{L}}$.

(ii) J is commutative, has neutral element 0 and absorbing element 1.

(iii) J is increasing (in each variable) if and only if I is a commutative quasi-copula.

(iv) If I is a commutative quasi-copula, then J also satisfies the 1-Lipschitz property.

Proof. Since the proofs of (i) and (ii) are trivial, we restrict ourselves to the proofs of (iii) and (iv).
(iii) Suppose I is a commutative quasi-copula and $x_1 \leq x_2$. Then,

$$\begin{aligned} & J(x_1, y) \leq J(x_2, y) \\ \Leftrightarrow\ & x_1 + y - I(x_1, y) \leq x_2 + y - I(x_2, y) \,. \end{aligned} \tag{3.6}$$

Invoking the 1-Lipschitz property of I, inequality (3.6) holds for any $y \in [0,1]$. Analogously, consider $y_1 \leq y_2$, then $J(x, y_1) \leq J(x, y_2)$ holds for any $x \in [0,1]$. In the same way, we can prove that if J is increasing, then I satisfies the 1-Lipschitz property.
(iv) Suppose I is a commutative quasi-copula and $(x_1, x_2, y_1, y_2) \in [0,1]^4$. Then, the following holds:

$$|J(x_1, y_1) - J(x_2, y_2)| \leq |J(x_1, y_1) - J(x_2, y_1)| + |J(x_2, y_1) - J(x_2, y_1)| .$$

Furthermore, suppose $x_1 \leq x_2$. Then,

$$\begin{aligned} |J(x_1, y_1) - J(x_2, y_1)| &= J(x_2, y_1) - J(x_1, y_1) \\ &= x_2 + y_1 - I(x_2, y_1) - x_1 - y_1 + I(x_1, y_1) \\ &\leq x_2 - x_1 \\ &= |x_1 - x_2| . \end{aligned}$$

Analogously, we can prove that $|J(x_1, y_1) - J(x_2, y_1)| \leq |x_1 - x_2|$ if $x_1 > x_2$. In the same way, the following holds:

$$|J(x_2, y_1) - J(x_2, y_1)| \leq |y_2 - y_1| ,$$

and therefore J fulfills the 1-Lipschitz property. □

The union of two fuzzy sets can be modelled by the operator J, as defined in Eq. 3.5. However, we can also model the union of two fuzzy sets such that the following identity, which holds in classical set theory,

$$|A \cup B| = |(A^c \cap B^c)^c| . \tag{3.7}$$

is preserved in the fuzzy case, as follows:

$$A \cup B(x) = 1 - I(1 - A(x), 1 - B(x)) = I^*(A(x), B(x)) .$$

If we want to preserve Eq. (3.4) as well as Eq. (3.7) in fuzzy set theory, the intersection of two fuzzy sets should be modelled by a stable commutative copula.

Proposition 3.5 *If J is increasing, then $J = I^*$ if and only if I is a stable commutative quasi-copula.*

Proof. Follows directly from the definition of a stable commutative quasi-copula. □

C. The identity $|A \setminus B| = |A| - |A \cap B|$

If we want to preserve the identity $|A \setminus B| = |A| - |A \cap B|$, then the fuzzy set difference should be defined by

$$A \setminus B(x) = A(x) - I(A(x), B(x))\,, \tag{3.8}$$

with I a commutative conjunctor. Based on the right-hand side of Eq. (3.8), we define the difference operator V by

$$V(x, y) = x - I(x, y)\,. \tag{3.9}$$

Proposition 3.6 *The operator V has the following properties:*

(i) V is a binary operation on $[0, 1]$ and $V(x, \cdot)$ is decreasing for any $x \in [0, 1]$.

(ii) $V(\cdot, y)$ is increasing for any $y \in [0, 1]$ if and only if I is a commutative quasi-copula.

Proof. Similar to the proof of Proposition 3.4. □

D. The identity $|A \triangle B| = |A \setminus B| + |B \setminus A|$

If we want to preserve the identity

$$|A \triangle B| = |A| + |B| - 2|A \cap B|\,,$$

then the symmetric difference of two fuzzy sets should be defined by

$$A \triangle B(x) = A(x) + B(x) - 2I(A(x), B(x))\,, \tag{3.10}$$

with I a commutative conjunctor. If the difference of fuzzy sets is defined by the operator V in Eq. (3.9), then it also holds that $|A \triangle B| = |A \setminus B| + |B \setminus A|$.

Remark that the identities $|A \setminus B| = |A| - |A \cap B|$ and $|A \triangle B| = |A \setminus B| + |B \setminus A|$ already have been treated in [19] using the Frank t-norm family.

3.4 Conclusions and indications for future research

In this chapter, we listed several possibilities to express the cardinality of a fuzzy set and have chosen to use the basic scalar cardinality throughout the remainder of this work. Finally, we provided the reader with a set of fuzzification rules to model the intersection, the union and the (symmetric) difference of two fuzzy sets.

This chapter is the ideal operating base for future research. In this work, we only consider the basic scalar cardinality to define the cardinality of a fuzzy set. We also model the union of two fuzzy sets by the operator J as defined in Eq. (3.5) such that the valuation property is fulfilled. Will the valuation property also be fulfilled for other definitions of cardinalities and how should the operator J then be defined ?

Also a more thorough analysis of the operator V to model the difference of two fuzzy sets and even more the operator to model the symmetric difference of two fuzzy sets creates a possibility for future research.

Chapter 4

The Bell inequalities

4.1 Introduction

In Chapter 5, we will show that the Bell inequalities are of particular interest in the context of cardinalities of fuzzy sets. The results on the fuzzified Bell inequalities, which will be expounded in this chapter, can be employed to formulate two meta-theorems that can be used to check the validity of more general inequalities on fuzzy cardinalities.

First of all, we give a historical overview of the Bell inequalities. Then, we describe all Bell inequalities concerning four events in which at most two events are intersected at the same time and rewrite them in the context of basic scalar cardinalities. In Section 4.5, we prove that some inequalities are fulfilled for (quasi-)copulas. Moreover, considering the Frank t-norm family, we identify all parameter values such that each of the Bell-type inequalities is fulfilled. We also study in detail the Bell-type inequalities for continuous t-norms. We prove that ordinal sums preserve the Bell-type inequalities, which was the motivation for studying continuous Archimedean t-norms only. Finally, for the most important parametric families of continuous Archimedean t-norms and each of the Bell-type inequalities, we identify the parameter values such that the corresponding t-norms satisfy the inequality considered.

4.2 Historical overview

The story of the Bell inequalities goes back to Einstein, Podolsky and Rosen. In 1935, they claimed that either quantum mechanics is incomplete, either we live in a world with ”spooky actions at a distance”, like Einstein said. Specifically, they showed that according to the theory one could put a particle in a measuring device at one location and, simply by doing that, instantly influence another particle arbitrarily far away (see Intermezzo 4.1). They refused to believe that this effect (known as non-locality) could really happen, and thus viewed it as evidence that quantum mechanics was incomplete. With their EPR–paradox, Einstein, Podolsky and Rosen tried to show that quantum mechanics isn't the final word. This proposition is known as the ”hidden variable theory”.

Intermezzo 4.1 The EPR–paradox

Consider the following quantum mechanical thought-experiment. Take a particle which is at rest and has spin zero. It spontaneously decays into two spin-$\frac{1}{2}$ particles (like electrons, protons, ...), which stream away in opposite directions at high speed. Due to the law of conservation of spin, we know that when one particle has spin up, the other particle has spin down. Which one is which? According to quantum mechanics, neither takes on a definite state until it is observed. The EPR–effect demonstrates that if one of the particles is detected, and its spin is measured, then the other particle, no matter where it is in the universe, instantaneously is forced to choose as well and take on the role of the other particle. So it seems that when you measure one of the particles you immediately measure the other, and this is an effect faster than light, so it goes against relativity, which says that nothing can travel faster than light. This is the EPR–paradox, which led Einstein to argue that quantum mechanics was not a complete theory.

The issue of the existence of hidden variables in quantum mechanics is almost as old as quantum mechanics itself. In 1964, Bell [4] showed that if one makes some ‘reasonable’ assumptions about the hidden variables, like locality and statistical independence of distant measurements, the correlations for the outcome of measurements for an Einstein–Podolsky–Rosen–like experiment have to satisfy a set of inequalities. When Bell introduced his inequalities, he had in mind the quantum mechanical situation originally introduced by Bohm [8] of correlated spin-$\frac{1}{2}$ particles in the singlet spin state.

Later on, Pitowsky [66] developed a generalization of the Bell inequalities where any number of experiments and events can be taken into account. He proved that the situation where the Bell inequalities are satisfied is equivalent to the situation where, for a set of probabilities connected to the outcomes of the considered experiments, there exists a Kolmogorovian probability model. This means that the probability P of some event E_i $(i = 1, \ldots, n)$ with respect to a finite universe X is defined in such a way that P must satisfy the three axioms of Kolmogorov:

(i) $0 \leq P(E_i) \leq 1$, for any $i = 1, \ldots, n$,

(ii) $P(X) = 1$,

(iii) Any countable sequence of pairwise disjoint events $E_1, \ldots, E_n$ satisfies $P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{i=1}^{n} P(E_i)$.

Pykacz and D'Hooghe [67] recently studied which of the numerous Bell-type inequalities that are necessarily satisfied by Kolmogorovian probabilities may be violated in various models of fuzzy probability calculus. They showed that the most popular model of fuzzy probability calculus based on minimum and maximum cannot be distinguished from the Kolmogorovian model by any of the inequalities studied by Pitowsky. They also proved that if one considers fuzzy set intersection pointwisely generated by a Frank t-norm $T_\lambda^{\mathbf{F}}$, then the borderline between models of fuzzy probability calculus that can be distinguished from Kolmogorovian ones and models that cannot be distinguished (by the same set of inequalities) is situated at $\lambda = 9 + 4\sqrt{5}$.

Nevertheless, the Bell-type inequalities are not only of interest to fuzzy probability calculus, they also appear in other applications of fuzzy logic. De Baets and De Meyer [19] proved that one particular inequality (inequality I_3^2 further on in this thesis) is of primordial importance in the design of transitivity-preserving fuzzification schemes for cardinality-based similarity measures. But, for more details concerning this we try the reader's patience to Chapter 6.

4.3 Bell-type inequalities in probability theory

Pitowsky [66] showed that the original Bell inequalities can be derived in a purely mathematical context without any reference to physics so

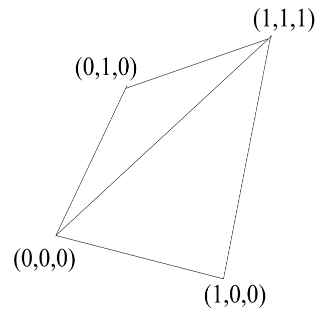


Figure 4.1: Geometrical interpretation of the Bell inequalities

that their range of applicability is by no way restricted to physical phenomena. He developed a generalization of the Bell inequalities and proved that the situation where the Bell inequalities are satisfied is equivalent to the situation where, for a set of probabilities connected to the outcomes of the considered experiments, there exists a (classical) Kolmogorovian probability model.

The probability of a single random event A_i is denoted by $p_i = P(A_i)$ and the probability of the intersection of a pair of random events A_i and A_j is denoted by $p_{ij} = P(A_i \cap A_j)$. Since p_i, p_j and p_{ij} are probabilities, the following inequalities hold:

$$\begin{aligned} p_{ij} &\leq \min(p_i, p_j)\,, \\ p_i + p_j - p_{ij} &\leq 1\,. \end{aligned}$$

They can be expressed jointly as the double inequality

$$T_{\mathbf{L}}(p_i, p_j) \leq p_{ij} \leq T_{\mathbf{M}}(p_i, p_j)\,.$$

Pitowsky also suggested a geometrical interpretation of these Bell inequalities [66].

Consider the three dimensional real space, and in it the set of all vectors of the form $p = (p_i, p_j, p_{ij})$, where p_i, p_j and p_{ij} satisfy the above Bell inequalities. This set is a closed convex polytope whose vertices are the extreme cases $(0,0,0)$, $(0,1,0)$, $(1,0,0)$ and $(1,1,1)$ (see Figure 4.1). Every convex polytope has a dual description: either in terms of its vertices or in terms of its facets. Under the first description, a given vector is an element of the polytope if and only if it can be represented as a convex combination of its vertices. Under the second description, a given vector is an element of the polytope if and only if its coordinates satisfy a set of linear inequalities which represent the supporting hyperplanes of the polytope. The existence for such a dual description for every polytope is known as the Minkowski-Weyl theorem.

In the specific case of the polytope in Figure 4.1 our starting point has been the second description. We can obtain the same result considering the first description.

Example 4.1 Consider two propositions:

a_1: it will rain in Brussels tomorrow,

a_2: it will rain in Ghent tomorrow.

Obviously, there are four possibilities that can be summed up in the following truth table:

a_1	a_2	a_1 and a_2
0	0	0
0	1	0
1	0	0
1	1	1

The rows in this table, if looked at as vectors in a three dimensional space, are just the vertices of our polytope. Suppose that we were to bet on each one of the four possibilities. Let λ_1 denote the probability we assign to the event "a_1 is false and a_2 is false", or equivalently, "it will not rain in Brussels nor in Ghent tomorrow" and so forth. Since there are only four possibilities and since they are mutually incompatible we must have $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$. Consider the vector

$$\begin{aligned}(p_1, p_2, p_{12}) &= \lambda_1(0,0,0) + \lambda_2(0,1,0) + \lambda_3(1,0,0) + \lambda_4(1,1,1) \\ &= (\lambda_3 + \lambda_4, \lambda_2 + \lambda_4, \lambda_4)\,.\end{aligned}$$

Since it is a convex combination of the vertices, (p_1, p_2, p_{12}) lies in the polytope.

Pitowsky proved the following theorem.

Theorem 4.1 [66] *The following statements are equivalent:*

1. *the numbers p_i, p_j and p_{ij} represent probabilities,*
2. *$p = (p_i, p_j, p_{ij})$ is a vector in the polytope,*
3. *the Bell inequalities are satisfied.*

In the case of experiments concerning three or four random events in which at most two events are intersected at the same time and only considering four possible intersections, Pitowsky found the following set of inequalities:

$$0 \leq p_i - p_{ij} - p_{ik} + p_{jk}\,, \tag{4.1}$$

$$p_i + p_j + p_k - p_{ij} - p_{ik} - p_{jk} \leq 1\,, \tag{4.2}$$

$$0 \leq p_i + p_k - p_{ik} - p_{il} - p_{jk} + p_{jl} \leq 1\,, \tag{4.3}$$

for any different i, j, k, l. Inequalities (4.1) and (4.2) are called the Bell–Wigner inequalities, while the double inequality (4.3) is referred to as the Clauser–Horne inequality.

Next to the above inequalities, we have generated the remaining Bell-type inequalities (taking into account five and six possible intersections) using the cdd package[1] of Fukuda [36]:

$$\begin{aligned}
0 \leq p_i + p_j + p_{ij} - p_{ik} - p_{il} - p_{jl} - p_{jk} + p_{kl}\,,\\
p_i + p_j + p_k + p_l - p_{ij} - p_{ik} - p_{il} - p_{jk} - p_{jl} - p_{kl} \leq 1\,,\\
2p_i + 2p_j + 2p_k + 2p_l - p_{ij} - p_{ik} - p_{il} - p_{jk} - p_{jl} - p_{kl} \leq 3\,,\\
0 \leq p_i - p_{ij} - p_{ik} - p_{il} + p_{jk} + p_{jl} + p_{kl}\,,\\
p_i + p_j + p_k - 2p_l - p_{ij} - p_{ik} + p_{il} - p_{jk} + p_{jl} + p_{kl} \leq 1\,,
\end{aligned}$$

for any different i, j, k, l.

4.4 Bell-type inequalities for commutative conjunctors

We can rewrite the above-mentioned Bell-type inequalities in the context of fuzzy probability calculus, or equivalent in the context of basic scalar cardinalities. Let A and B be two fuzzy sets in a finite universe X of cardinality n. Although there are several possible means to define fuzzy probabilities, we adopt here the definition as introduced by Zadeh [99]: $P(A) = \sum_u A(u)/n$.

For instance, the classical inequality

$$P(A) + P(B) - P(A \cap B) \leq 1$$

[1] The cdd package is a C++ implementation of the "double description method" of Motzkin et al. [60] for generating all vertices and extreme rays of a general convex polyhedron in $\mathbb{R}^d$ given by a system of linear inequalities: $P = \{x \mid Ax \leq b\}$, where A is an $m \times d$ real matrix and b is a real m-dimensional vector, see http://www.cs.mcgill.ca/~fukuda/soft/cdd_home/cdd.html

can be expressed for fuzzy probabilities, with $A \cap B$ pointwisely modelled by means of a commutative conjunctor I, in the following way,

$$\frac{\sum_u A(u)}{n} + \frac{\sum_u B(u)}{n} - \frac{\sum_u I(A(u), B(u))}{n} \leq 1 \,. \tag{4.4}$$

The latter inequality is fulfilled when

$$A(u) + B(u) - I(A(u), B(u)) \leq 1$$

for any $u \in X$, which in turn is fulfilled when

$$x + y - I(x, y) \leq 1 \,, \tag{4.5}$$

for any $(x, y) \in [0, 1]^2$. Hence, (4.5) is a sufficient condition for (4.4). On the other hand, considering (4.4) for $n = 1$ and arbitrary A and B (which are then fuzzy singletons in a one-point universe), then (4.4) is equivalent to (4.5).

In general, the Bell-type inequalities for commutative conjunctors are necessary and sufficient conditions for the corresponding Bell-type inequalities for fuzzy probabilities to hold for any fuzzy sets in a universe of cardinality n. This leads to the Bell-type inequalities listed in Table 4.1. To simplify the discussion of these inequalities we introduce a unique code I_i^j for each inequality where i denotes the number of events involved and j is a serial number.

Remark 4.1 *Note that the double inequality I_2^1 is superfluous since it can be obtained from I_3^2 by putting $x = 1$, resp. $z = 0$. In general, considering a higher number of events renders the previous level obsolete. Inequality I_3^2 can be obtained by putting $t = 0$ in inequalities I_4^5 or I_4^8. Putting $t = 0$ in inequality I_4^6 or I_4^9, we obtain inequality I_3^3.*

When considering more general scalar cardinalities, in particular using as scaling function an automorphism ϕ of the unit interval, the validity of the Bell-type inequalities for these more general scalar cardinalities is equivalent to the validity of the inequalities I_2^1–I_4^9 applied to the ϕ^{-1}-transform of I, *i.e.* the commutative conjunctor $I^{\phi^{-1}}$ defined by $I^{\phi^{-1}} = \phi\left(I(\phi^{-1}(x), \phi^{-1}(y))\right)$. As this does not bring any additional insights and would only make notations heavier, we will restrict ourselves to basic scalar cardinalities from here on.

Code	Bell-type inequalities
I_2^1	$T_{\mathbf{L}} \leq I \leq T_{\mathbf{M}}$
I_3^2	$0 \leq x - I(x,y) - I(x,z) + I(y,z)$
I_3^3	$x + y + z - I(x,y) - I(x,z) - I(y,z) \leq 1$
I_4^4	$0 \leq x + t - I(x,z) - I(x,t) - I(y,t) + I(y,z) \leq 1$
I_4^5	$0 \leq x + t - I(x,y) - I(x,z) + I(x,t) + I(y,z) - I(y,t) - I(z,t)$
I_4^6	$x + y + z + t - I(x,y) - I(x,z) - I(x,t) - I(y,z) - I(y,t) - I(z,t) \leq 1$
I_4^7	$2x + 2y + 2z + 2t - I(x,y) - I(x,z) - I(x,t) - I(y,z) - I(y,t) - I(z,t) \leq 3$
I_4^8	$0 \leq x - I(x,y) - I(x,z) - I(x,t) + I(y,z) + I(y,t) + I(z,t)$
I_4^9	$x + y + z - 2t - I(x,y) - I(x,z) + I(x,t) - I(y,z) + I(y,t) + I(z,t) \leq 1$

Table 4.1: Bell-type inequalities for commutative conjunctors.

4.5 Bell-type inequalities for quasi-copulas and copulas

As we were interested in modelling the intersection of two fuzzy sets by a commutative conjunctor in the previous section, we now focus our attention on three particular classes of conjunctors: the class of triangular norms (t-norms), the class of copulas and the class of quasi-copulas.

We will show that inequalities I_2^1, I_3^2 and I_4^4, and some generalization of I_3^2 and I_4^4 are fulfilled for any quasi-copula, while for inequality I_4^5 the class of quasi-copulas needs to be further restricted to the subclass of copulas. Indeed, commutativity does not enter the play here, rendering these inequalities available for a broader context as well. For the remaining inequalities, it will follow from our study of these inequalities for the Frank t-norm family in the next section, that they are not generally valid for commutative copulas.

Theorem 4.2 *Inequalities I_2^1, I_3^2 and I_4^4 are fulfilled for any quasi-copula C.*

Proof. The double inequality I_2^1 is nothing else but Proposition 3.1. Next, we prove that I_3^2 is satisfied, i.e.

$$0 \leq x - C(x,y) - C(x,z) + C(y,z)\,. \tag{4.6}$$

We distinguish two cases:

(i) In case $x \leq y$, we have that $C(x,z) \leq C(y,z)$. Since $C(x,y) \leq x$, (4.6) trivially holds.

(ii) In case $x > y$, the 1-Lipschitz property implies that

$$0 \leq x - y - C(x,z) + C(y,z)\,.$$

Since $C(x,y) \leq y$, (4.6) again holds.

Finally, we prove that the double inequality I_4^4, i.e.

$$0 \leq x + t - C(x,z) - C(x,t) - C(y,t) + C(y,z) \leq 1 \tag{4.7}$$

holds. The left-hand side of (4.7) is obtained through a double application of inequality I_3^2:

$$\begin{aligned}
&x + t - C(x,z) - C(x,t) - C(y,t) + C(y,z)\\
&\geq x - C(x,z) - C(x,y) + C(y,z)\\
&\geq 0\,.
\end{aligned}$$

Next, we prove the right-hand side. Again, we distinguish two cases:

(i) In case $z \leq t$, it follows using $x + t - C(x,t) \leq 1$ and $C(y,z) \leq C(y,t)$ that

$$\begin{aligned} & x + t - C(x,z) - C(x,t) - C(y,t) + C(y,z) \\ & \leq 1 - C(x,z) - C(y,t) + C(y,z) \\ & \leq 1\,. \end{aligned}$$

(ii) In case $z > t$, it follows using $x + z - C(x,z) \leq 1$ that

$$\begin{aligned} & x + t - C(x,z) - C(x,t) - C(y,t) + C(y,z) \\ & = x + t + z - z - C(x,z) - C(x,t) - C(y,t) + C(y,z) \\ & \leq 1 + t - z - C(x,t) - C(y,t) + C(y,z) \\ & \leq 1 + t - z - C(y,t) + C(y,z)\,. \end{aligned}$$

Invoking the 1-Lipschitz property of C, we can conclude that the right-hand side of (4.7) holds.

□

The following example shows that inequality I_3^2 is not characteristic for the class of quasi-copulas.

Example 4.2 Consider the following commutative conjunctor [54] (see Figure 4.2)

$$K(x,y) = \max(x + y - 1, 0)\,(2 - \max(x,y))\,, \tag{4.8}$$

which is bounded from below by $T_{\mathbf{L}}$. This operator does not fulfil the 1-Lipschitz property. Consider $x_1 = 0.1, x_2 = 0.6$ and $y_1 = y_2 = 0.9$, then

$$|K(x_1,y_1) - K(x_2,y_2)| = 0.55 > 0.5 = |x_1 - x_2| + |y_1 - y_2|\,.$$

Nevertheless, this operator satisfies inequality I_3^2, i.e.

$$0 \leq x - K(x,y) - K(x,z) + K(y,z)\,.$$

Indeed, since K is monotone and $K \leq T_{\mathbf{M}}$, the only non-trivial case turns out to be $x > \max(y,z)$. In view of symmetry, we consider for instance the case $x > y \geq z$ and verify that

$$\begin{aligned} 0 \leq\ & x - \max(x + y - 1, 0)(2 - x) - \max(x + z - 1, 0)(2 - x) \\ & + \max(y + z - 1, 0)(2 - y)\,. \end{aligned}$$

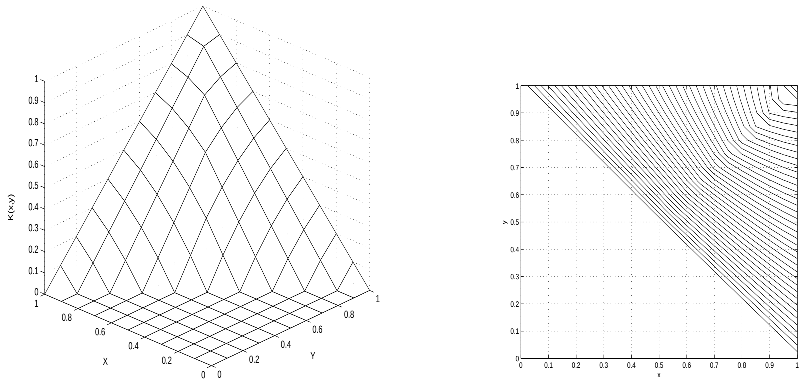


Figure 4.2: 3D-plot and contour plot of the operator K.

Consider for instance the case $x+y \geq 1$, $x+z \geq 1$ and $y+z \geq 1$, the other ones being similar. It is easily verified that the above inequality is equivalent to

$$0 \leq 2x^2 - y^2 + xy + xz - yz - 5x + y + 2\,,$$

or also

$$0 \leq (y+z-1)(x-y) + 2(1-x)^2\,,$$

which is clearly fulfilled.

Theorem 4.3 *For any conjunctor I that satisfies inequalities I_3^2 and I_4^4, the following inequality holds for any $n \geq 3$:*

$$0 \leq \sum_{i=2}^{n-1} x_i - \sum_{i=1}^{n-1} I(x_i, x_{i+1}) + I(x_1, x_n) \leq \left\lceil \frac{n}{2} \right\rceil - 1 \tag{4.9}$$

for any $(x_1, \ldots, x_n) \in [0,1]^n$.

Proof. In case $n = 3$, the double inequality (4.9) reduces to

$$0 \leq x_2 - I(x_1,x_2) - I(x_2,x_3) + I(x_1,x_3) \leq 1\,.$$

The left-hand side being inequality I_3^2, we only need to prove the right-hand side. As observed in Remark 4.1, inequality I_3^2 implies in particular inequality I_2^1. Hence, it indeed holds that

$$\begin{aligned}
&x_2 - I(x_1,x_2) - I(x_2,x_3) + I(x_1,x_3)\\
&\leq 1 - x_3 - I(x_1,x_2) + I(x_1,x_3)\\
&\leq 1\,.
\end{aligned}$$

The case $n = 4$ is nothing else but inequality I_4^4.

Now let $n \geq 5$. Suppose that the left-hand side of (4.9) holds for $n-1$, then we use induction and inequality I_3^2 to prove that it also holds for n:

$$\begin{aligned}
&x_2 + \ldots + x_{n-1} - I(x_1, x_2) - \ldots - I(x_{n-1}, x_n) + I(x_1, x_n) \\
&= (x_2 + \ldots + x_{n-2}) + x_{n-1} - (I(x_1, x_2) + \ldots + I(x_{n-2}, x_{n-1})) \\
&\quad - I(x_{n-1}, x_n) + I(x_1, x_n) \\
&\geq x_{n-1} - I(x_1, x_{n-1}) - I(x_{n-1}, x_n) + I(x_1, x_n) \\
&\geq 0\,.
\end{aligned}$$

Next, suppose the right-hand side of (4.9) holds for $n-2$, then we use induction and I_4^4 to prove:

$$\begin{aligned}
&x_2 + \ldots + x_{n-1} - I(x_1, x_2) - \ldots - I(x_{n-1}, x_n) + I(x_1, x_n) \\
&\leq \left\lceil \frac{n-2}{2} \right\rceil - 1 + x_{n-2} + x_{n-1} \\
&\quad - I(x_1, x_{n-2}) - I(x_{n-2}, x_{n-1}) - I(x_{n-1}, x_n) + I(x_1, x_n) \\
&\leq \left\lceil \frac{n-2}{2} \right\rceil - 1 + 1 = \left\lceil \frac{n}{2} \right\rceil - 1\,.
\end{aligned}$$

Since (4.9) holds for $n = 3$ and $n = 4$ and induction is either based on $n-2$ or $n-1$, this completes the proof. □

Theorem 4.4 *Inequality I_4^5 is fulfilled for any copula C.*

Proof. In the case $x \leq y$, $t \leq z$, the moderate growth property implies that

$$0 \leq -C(x, z) + C(x, t) + C(y, z) - C(y, t)\,.$$

Since also $0 \leq x - C(x, y)$ and $0 \leq t - C(z, t)$, a simple addition leads to

$$0 \leq x + t - C(x, y) - C(x, z) + C(x, t) + C(y, z) - C(y, t) - C(z, t)\,.$$

The proof of the remaining cases is analogous. □

The following example illustrates that inequality I_4^5 is not fulfilled for all commutative quasi-copulas.

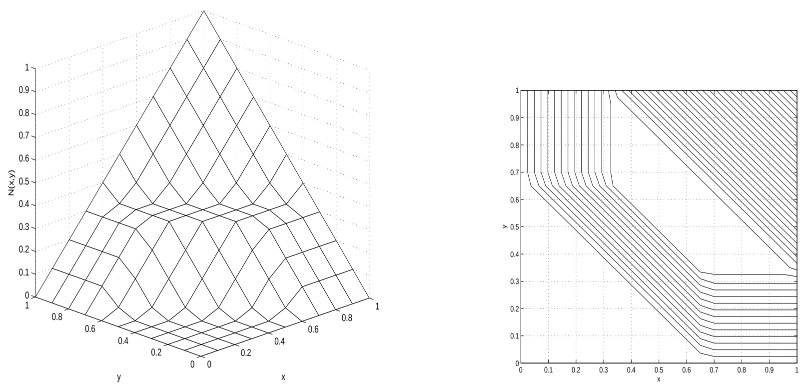


Figure 4.3: 3D-plot and contour plot of the operator N (4.10).

Example 4.3 Consider the binary operator N defined by

$$N(x,y) = \begin{cases} \min(x, y, \frac{1}{3}, x+y-\frac{2}{3}) & \text{if } \frac{2}{3} \leq x+y \leq \frac{4}{3}\,, \\ \max(x+y-1, 0) & \text{otherwise}\,, \end{cases} \tag{4.10}$$

which is a commutative quasi-copula, but not a copula [61, p. 13]. However, it does not satisfy inequality I_4^5. Consider $x = 0.4$, $y = z = 0.7$ and $t = 0.3$, then

$$x + t - N(x,y) - N(x,z) + N(x,t) + N(y,z) - N(y,t) - N(z,t)$$

evaluates to -0.13 and is strictly negative.

4.6 Bell-type inequalities for Frank t-norms

Pykacz and D'Hooghe [67] demonstrated that a set of Bell-type inequalities (I_2^1, I_3^2, I_3^3 and I_4^4) does not allow to distinguish Kolmogorovian probabilities from fuzzy probabilities based on minimum and maximum, but it allows to distinguish all these models from fuzzy probability models based on the Łukasiewicz t-norm. They proved that if one uses fuzzy set intersections pointwisely generated by a Frank t-norm $T_\lambda^{\mathbf{F}}$, then the borderline between fuzzy probability models that can be distinguished from Kolmogorovian ones and these that cannot be distinguished is situated at $\lambda = 9 + 4\sqrt{5}$. In this section, we want to complete the results of Pykacz and D'Hooghe.

To be perfectly clear, we will recall the Frank t-norm family $\{T^{\mathbf{F}}_\lambda\}_{\lambda\in[0,\infty]}$. This family is given by

$$T^{\mathbf{F}}_\lambda(x,y)=\begin{cases} T_{\mathbf{M}}(x,y) & \text{if } \lambda=0\,,\\ T_{\mathbf{P}}(x,y) & \text{if } \lambda=1\,,\\ T_{\mathbf{L}}(x,y) & \text{if } \lambda=\infty\,,\\ \log_\lambda\left(1+\frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}\right) & \text{otherwise}\,.\end{cases}$$

Since all Frank t-norms are copulas, inequalities I_2^1, I_3^2, I_4^4 and I_4^5 obviously hold for all Frank t-norms. Now it only remains to identify the parameter values for which inequalities I_3^3, I_4^6, I_4^7, I_4^8 and I_4^9 are fulfilled. The results of this study are summarized in Table 4.2.

4.6.1 Inequality I_3^3

Let us define a function f as follows:

$$f(x,y,z)=x+y+z-T(x,y)-T(x,z)-T(y,z)-1\,. \tag{4.11}$$

Writing the right-hand side of (4.11) explicitly for a Frank t-norm with $\lambda\in]0,1[\cup]1,\infty[$, we obtain

$$\begin{aligned} f(x,y,z)\,=\,& x+y+z-\log_\lambda\left(1+\frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}\right)\\ &-\log_\lambda\left(1+\frac{(\lambda^x-1)(\lambda^z-1)}{\lambda-1}\right)\\ &-\log_\lambda\left(1+\frac{(\lambda^y-1)(\lambda^z-1)}{\lambda-1}\right)-1\,.\end{aligned}$$

In order to find the stationary points of f, we set the first-order derivatives of f equal to zero, and obtain:

$$\begin{aligned} f_x(x,y,z)\,&=\,0\,,\\ f_y(x,y,z)\,&=\,0\,,\\ f_z(x,y,z)\,&=\,0\,,\end{aligned}$$

or equivalently,

$$1-\frac{\lambda^x(\lambda^y-1)}{\lambda+\lambda^{x+y}-\lambda^x-\lambda^y}-\frac{\lambda^x(\lambda^z-1)}{\lambda+\lambda^{x+z}-\lambda^x-\lambda^z}\,=\,0\,, \tag{4.12}$$

$$1-\frac{\lambda^y(\lambda^x-1)}{\lambda+\lambda^{x+y}-\lambda^x-\lambda^y}-\frac{\lambda^y(\lambda^z-1)}{\lambda+\lambda^{y+z}-\lambda^y-\lambda^z}\,=\,0\,, \tag{4.13}$$

$$1-\frac{\lambda^z(\lambda^x-1)}{\lambda+\lambda^{x+z}-\lambda^x-\lambda^z}-\frac{\lambda^z(\lambda^y-1)}{\lambda+\lambda^{y+z}-\lambda^y-\lambda^z}\,=\,0\,. \tag{4.14}$$

Subtracting (4.13) from (4.12) yields the following:

$$\begin{aligned}
&\frac{\lambda^x - \lambda^y}{\lambda + \lambda^{x+y} - \lambda^x - \lambda^y} \\
&\quad -(\lambda^z - 1)\left(\frac{\lambda^x}{\lambda + \lambda^{x+z} - \lambda^x - \lambda^z} - \frac{\lambda^y}{\lambda + \lambda^{y+z} - \lambda^y - \lambda^z}\right) = 0 \\
\Leftrightarrow\ &(\lambda^x - \lambda^y)\left(\frac{1}{\lambda + \lambda^{x+y} - \lambda^x - \lambda^y}\right. \\
&\quad \left. - \frac{(\lambda^z - 1)(\lambda - \lambda^z)}{(\lambda + \lambda^{x+z} - \lambda^x - \lambda^z)(\lambda + \lambda^{y+z} - \lambda^y - \lambda^z)}\right) = 0 \\
\Leftrightarrow\ &\lambda^x = \lambda^y \\
\Leftrightarrow\ &x = y\,.
\end{aligned}$$

In the same way, subtracting (4.14) from (4.12) renders $x = z$. Inside the unit cube $[0,1]^3$ the first-order derivatives of the function f can only be zero in the symmetric case $x = y = z$. Therefore, to identify the parameter values for which inequality I_3^3 is satisfied, we can investigate the inequality

$$3x - 3T(x,x) \leq 1. \tag{4.15}$$

Note that this inequality is trivially fulfilled for $T_{\mathbf{M}}$ ($\lambda = 0$) and $T_{\mathbf{P}}$ ($\lambda = 1$). Pykacz and D'Hooghe [67] have proven that inequality (4.15) is fulfilled for any Frank t-norm $T_\lambda^{\mathbf{F}}$, with $\lambda \leq 9+4\sqrt{5}$ (for instance, I_3^3 is not fulfilled for $T_{\mathbf{L}}$). From this, we can also conclude that inequality I_3^3 does not hold for all commutative copulas.

Pykacz and D'Hooghe also provide a nice example of a physical realization where inequality I_3^3 is violated [67].

Example 4.4 The experiment consists of pouring water into a "black box" and then checking whether it leaks through its bottom. The result of the experiment is positive if the floor under the box remains dry and negative otherwise. Inside the box there is a cylindrical vessel V of capacity 1 liter and a gun aimed at it. We assume that the (Kolmogorovian) probability that a bullet shot from the gun makes a hole at a specific height of the cylinder V is uniformly distributed along the height of V. Random experiments that we are going to perform consist of a "public" part: pouring water into the box, and a "hidden" part: shooting the gun once. Since the "hidden" part can be kept in secret, all the experimenter can do is to pour water into the box and check weather it leaks. If one pours $0 \leq q \leq 1$ of a liter of water into the box, it leaves the upper $1 - q$ part of V empty, so the probability that the floor remains dry is $P = 1 - q$. Of course, if one pours more than 1 liter, water will surely pour out of V, so $P = 0$.

Let us prepare vessels V_i (with $i = 1, 2, \ldots, n$), containing (for any i) $1 - p_i$ of a liter of water and let experiment E_i consist of pouring water from a vessel V_i into the box and checking whether the floor under the box remains dry. According to the previous considerations $P = p_i$, for any i. Let us also consider the conjunction of experiments E_i and E_j consisting of the simultaneous pouring of water from vessels V_i and V_j into the box and subsequent checking whether the floor remains dry.

If $1 - p_i + 1 - p_j \geq 1$, i.e. if $p_i + p_j \leq 1$, then water obviously pours out onto the floor so $p_{ij} = P(E_i \cap E_j) = 0$. However, if $p_i + p_j \geq 1$, then the fraction $p_i + p_j - 1$ of the cylinder remains empty, so $p_{ij} = P(E_i \cap E_j) = p_i + p_j - 1$. Therefore, we see that probabilities of conjunctions of random events E_i have to be calculated using the Łukasiewicz t-norm and Bell-type inequality I_3^3 may be violated. For example, for $p_1 = p_2 = p_3 = 0.5$, we get $p_1 + p_2 + p_3 - p_{12} - p_{13} - p_{23} = 0.5 + 0.5 + 0.5 - 0 - 0 - 0 = 1.5$.

4.6.2 Inequalities I_4^6 and I_4^7

In this subsection, we want to give a detailed description of the calculations made to obtain the parameter values for a Frank t-norm T such that inequalities I_4^6 and I_4^7 are satisfied. We consider first inequality I_4^6. Let

$$\begin{aligned} f(x,y,z,t) &= x + y + z + t - T(x,y) - T(x,z) - T(x,t) \\ &\quad - T(y,z) - T(y,t) - T(z,t)\,. \end{aligned}$$

It can easily be shown that inside the unit hypercube $[0,1]^4$, the first-order derivatives of f, with T belonging to the Frank t-norm family, can only be zero in the symmetric case $x = y = z = t$. Inequality I_4^6 is then equivalent to

$$4x - 6T(x,x) \leq 1\,. \tag{4.16}$$

This inequality is trivially fulfilled for $T_{\mathbf{M}}$ ($\lambda = 0$) and $T_{\mathbf{P}}$ ($\lambda = 1$). Writing inequality (4.16) explicitly for $\lambda \in]0,1[\cup]1,\infty[$, we obtain

$$4x - 6\log_\lambda\left(1 + \frac{(\lambda^x - 1)^2}{\lambda - 1}\right) \leq 1\,.$$

Let us define a function g as follows:

$$g(x) = 4x - 6\log_\lambda\left(1 + \frac{(\lambda^x - 1)^2}{\lambda - 1}\right)$$

and investigate for any fixed λ what is the maximum g attains in $[0, 1]$. The first-order derivative of g equals

$$g'(x) = 4 - \frac{12(\lambda^x - 1)\lambda^x}{\lambda^x + \lambda - 2}\,.$$

Solving $g'(x) = 0$, we find that g reaches an extremal value in $x_s = \log_\lambda\left(\frac{1+\sqrt{1+8\lambda}}{4}\right)$ which can be verified to be located in $]0, 1[$. Straightforward computation yields $g''(x_s) < 0$ and g attains its maximum in x_s. Let us determine for which values of λ this maximum $g(x_s)$ equals 1, i.e. we solve $g(x_s) = 1$ for λ. Using Maple, we find the unique solution $\lambda_s = 9.29469$. Since inequality (4.16) holds for $\lambda = 0$ and $\lambda = 1$, we can conclude that inequality I_4^6 holds for any $\lambda \in [0, 9.29469]$.

We have already noticed that inequality I_3^3 can be obtained from inequality I_4^6 by putting $t = 0$, and indeed $9.29469 \leq 9 + 4\sqrt{5}$.

The same reasoning can be made to see that inequality I_4^7, for all members of the Frank t-norm family, is equivalent to

$$8x - 6T(x, x) \leq 3\,. \tag{4.17}$$

The parameter values for which inequality (4.17) is fulfilled, can also be obtained in the same way. Surprisingly, the same upper bound $\lambda_s = 9.29469$ is obtained.

4.6.3 Inequalities I_4^8 and I_4^9

Also for inequalities I_4^8 and I_4^9, we want to give a detailed overview of the calculations made in order to obtain the parameter values. Let

$$\begin{aligned} f(x, y, z, t) = {} & -x + T(x, y) + T(x, z) + T(x, t) \\ & -T(y, z) - T(y, t) - T(z, t)\,. \end{aligned}$$

Again, it can easily be shown that inside the unit hypercube $[0, 1]^4$, the first-order derivatives of f, with T belonging to the Frank t-norm family, can only be zero in the symmetric case $y = z = t$. Inequality I_4^8 is then equivalent to

$$-x + 3T(x, y) - 3T(y, y) \leq 0\,. \tag{4.18}$$

This inequality is trivially fulfilled for $T_{\mathbf{M}}$ ($\lambda = 0$). Writing inequality (4.18) explicitly for $\lambda \in]0, 1[\cup]1, \infty[$, we obtain

$$-x + 3\,\log_\lambda\left(1 + \frac{(\lambda^x - 1)(\lambda^y - 1)}{\lambda - 1}\right) - 3\,\log_\lambda\left(1 + \frac{(\lambda^y - 1)^2}{\lambda - 1}\right) \leq 0\,.$$

Let us define a function g as follows:

$$g(x,y) = -x + 3\,\log_\lambda\left(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}\right) - 3\,\log_\lambda\left(1+\frac{(\lambda^y-1)^2}{\lambda-1}\right)$$

and determine the values of λ for which $g(x,y) \leq 0$ for all $(x,y) \in [0,1]^2$. In order to find the stationary points of g, we set the first-order derivatives of g equal to zero, and obtain:

$$\begin{aligned}
g_x(x,y) &= 3\,\frac{\lambda^x(\lambda^y-1)}{\lambda+\lambda^{x+y}-\lambda^x-\lambda^y} = 0\,,\\
g_y(x,y) &= 3\,\frac{\lambda^y(\lambda^x-1)}{\lambda+\lambda^{x+y}-\lambda^x-\lambda^y} - \,6\,\frac{\lambda^y(\lambda^y-1)}{\lambda+\lambda^{2y}-2\lambda^y} = 0\,.
\end{aligned}$$

Introducing the short notations $s=\lambda^x$ and $t=\lambda^y$, we solve the above equations for λ and t, and obtain:

$$\lambda = \frac{2s(2s^2+1)}{5s+1} = 1 + \frac{(2s-1)(s-1)}{5s+1}\,, \tag{4.19}$$

$$t = \frac{2s(s+2)}{5s+1} = 1 + \frac{(2s+1)(s-1)}{5s+1}\,, \tag{4.20}$$

from which it is easily verified that $\lambda, t \in \,]0,1[$ when $s \in \,]0,1[$ and $\lambda, t \in \,]1,\infty[$ when $s \in \,]1,\infty[$. Also notice that $\lambda < s$ when $\lambda < 1$ and $\lambda > s$ when $\lambda > 1$. It follows that the function g has a stationary point (x_s, y_s) located in $[0,1]^2$ for all $\lambda \in \,]0,1[\,\cup\,[1,\infty[$. Furthermore, using (4.19) and (4.20), we compute the value $g(x_s,y_s)$ in that stationary point is given by

$$g(x_s,y_s) = \log_\lambda\left(\frac{(5s+1)^3}{8s(2s+1)^3}\right)\,.$$

For $s \in \,]0,1[$, it holds that $(5s+1)^3 > 8s(2s+1)^2$ and since in that case $\lambda \in \,]0,1[$, it follows that $g(x_s,y_s) < 0$, whereas for $s \in \,]1,\infty[$, it holds that $(5s+1)^3 < 8s(2s+1)^2$ and since $\lambda \in \,]1,\infty[$, it again follows that $g(x_s,y_s) < 0$. However, we still need to verify whether the stationary point (x_s,y_s) is a maximum, a minimum or a saddle point of g. To that end, we compute the second order derivatives of g. Using (4.19) and (4.20), we obtain:

$$\begin{aligned}
g_{xx}(x_s,y_s) &= \frac{2}{3}\,\ln\lambda\,,\\
g_{xy}(x_s,y_s) &= \frac{2}{3}\left(\frac{s+2}{s-1}\right)\ln\lambda\,,\\
g_{yy}(x_s,y_s) &= -\,4\left(\frac{s(s+2)^2}{(s-1)(2s+1)^2}\right)\ln\lambda\,,
\end{aligned}$$

Code	Values for λ
I_2^1	$\lambda \in \mathbb{R}^+$
I_3^2	$\lambda \in \mathbb{R}^+$
I_3^3	$\lambda \in [0, 9 + 4\sqrt{5}]$
I_4^4	$\lambda \in \mathbb{R}^+$
I_4^5	$\lambda \in \mathbb{R}^+$
I_4^6	$\lambda \in [0, 9.29]$
I_4^7	$\lambda \in [0, 9.29]$
I_4^8	$\lambda \in [0, 9 + 4\sqrt{5}]$
I_4^9	$\lambda \in [0, 9 + 4\sqrt{5}]$

Table 4.2: Values of the parameter λ for which the Bell-type inequalities are fulfilled considering the Frank t-norm family $\{T_\lambda^{\mathbf{F}}\}_{\lambda \in [0,\infty]}$.

from which it follows that the determinant of the second-order derivatives of g in the stationary point (x_s, y_s) is given by

$$A_2 = [g_{xx} g_{yy} - {g_{xy}}^2](x_s, y_s) = \frac{-4(s+2)^2(110s^2 + 106s + 27)}{3(s-1)^2(2s+1)^2} (\ln \lambda)^2 \, .$$

Since $A_2 < 0$ for all $s \in]0,1[\cup]1, \infty[$, the stationary point (x_s, y_s) is neither a minimum, nor a maximum (it is a saddle point), and therefore the maximum of g is reached on the boundary of the domain $[0,1]^2$. It is easy to see that on the edges $x = 0, y = 0, y = 1$, inequality (4.18) is always fulfilled, while the edge $x = 1$ leads to inequality (4.28), which is only fulfilled for all $y \in [0,1]$ if $\lambda \leq 9 + 4\sqrt{5}$.

The case $\lambda = 1$ can be verified separately in a similar manner. Hence, we can conclude that inequality I_4^8 holds for any $\lambda \in [0, 9{+}4\sqrt{5}]$.

The same reasoning can be made to see that inequality I_4^9, for all members of the Frank t-norm family, is equivalent to

$$3x - 2y - 3T(x,x) + 3T(x,y) \leq 1 \, . \tag{4.21}$$

The values of λ such that inequality (4.21) is fulfilled, can also be obtained in a similar way. Again, the same upper bound $\lambda = 9 + 4\sqrt{5}$ is obtained.

4.7 Bell-type inequalities for ordinal sums

Ling [58] has shown that for every continuous t-norm T, either $T = T_{\mathbf{M}}$, T is Archimedean or T is the ordinal sum of a family of continuous Archimedean t-norms. We clarify the notions used in this statement.

Definition 4.1 [52] *A t-norm T is called Archimedean if*

$$(\forall (x,y) \in]0,1[^2)(\exists n \in \mathbb{N})(x_T^{(n)} < y)\,,$$

with $x_T^{(0)} = 1$ and $x_T^{(n)} = T(x, x_T^{(n-1)})$.

As the t-norms considered in this thesis are always continuous, it suffices to know that

Proposition 4.1 [52] *A continuous t-norm T is Archimedean if and only if*

$$(\forall x \in]0,1[)(T(x,x) < x)\,.$$

Note that $T_{\mathbf{M}}$ is not Archimedean, while $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ are.

This section consists of a single theorem stating that ordinal sums preserve Bell-type inequalities. We conclude the section with a more general conjecture.

Theorem 4.5 *Consider any of the Bell-type inequalities. The ordinal sum of a family of t-norms fulfils this inequality if and only if each of the summands fulfils this inequality.*

Proof. The fact that the summands of an ordinal sum fulfil a given Bell-type inequality when the ordinal sum does, is easily verified. The converse is more tedious. Unfortunately, at this moment, each of the inequalities requires its own proof. To illustrate the line of reasoning, we consider for instance inequalities I_3^3 and I_4^4. The proofs for the other inequalities are similar and are mainly case-based.

Inequality I_3^3. First we remark that substituting $x = 0$ in I_3^3 yields the left part of I_2^1, i.e. $T_{\mathbf{L}} \leq T$. Now let T be the ordinal sum of a family of t-norms that fulfil I_3^3. Due to the symmetry of I_3^3 in x, y and z, we can assume without loss of generality that $x \leq y \leq z$. If x and y, as well as y and z, do not belong to same summand, then I_3^3 is fulfilled since it holds for $T_{\mathbf{M}}$.

(i) If x and z belong to the same summand $\langle a, b, T^*\rangle$, with T^* a t-norm that fulfils inequality I_3^3, then also y belongs to this summand. We can rewrite I_3^3 as follows

$$x + y + z - (a + (b-a)T^*(x', y')) - (a + (b-a)T^*(x', z')) \\ - (a + (b-a)T^*(y', z')) \leq 1\,,$$

with $x' = \frac{x-a}{b-a}$, $y' = \frac{y-a}{b-a}$ and $z' = \frac{z-a}{b-a}$. The latter inequality is equivalent to

$$x' + y' + z' - T^*(x', y') - T^*(x', z') - T^*(y', z') \leq \frac{1}{b-a}\,.$$

Since I_3^3 holds for T^* and $1 \leq \frac{1}{b-a}$, the above also holds.

(ii) If x and z do not belong to the same summand, i.e. $T(x, z) = T_{\mathbf{M}}(x, z) = x$, then we have to prove the following inequality:

$$y + z - T(x, y) - T(y, z) \leq 1\,. \tag{4.22}$$

(a) If x and y belong to the same summand $\langle a, b, T^*\rangle$, then $T(y, z) = T_{\mathbf{M}}(y, z) = y$ and (4.22) is equivalent to

$$\frac{z-a}{b-a} - T^*(x', y') \leq \frac{1}{b-a}\,,$$

with $x' = \frac{x-a}{b-a}$ and $y' = \frac{y-a}{b-a}$. It easily follows that

$$\frac{z-a}{b-a} - T^*(x', y') \leq \frac{1}{b-a} - T^*(x', y') \leq \frac{1}{b-a}\,.$$

(b) If y and z belong to the same summand $\langle a, b, T^*\rangle$, then $T(x, y) = T_{\mathbf{M}}(x, y) = x$ and (4.22) is equivalent to

$$y' + z' - \frac{x-a}{b-a} - T^*(y', z') \leq \frac{1}{b-a}\,,$$

with $y' = \frac{y-a}{b-a}$ and $z' = \frac{z-a}{b-a}$. Since T^* fulfils I_3^3, it holds that $T_{\mathbf{L}} \leq T^*$ and in particular $y' + z' - T^*(y', z') \leq 1$. It then follows that

$$y' + z' - \frac{x-a}{b-a} - T^*(y', z') \leq 1 - \frac{x-a}{b-a} = \frac{b-x}{b-a} \leq \frac{1}{b-a}\,.$$

Inequality I_4^4. First we remark that substituting $y = z = 0$ in I_4^4 again yields the left part of I_2^1, i.e. $T_{\mathbf{L}} \leq T$. Now let T be the ordinal sum of a family of t-norms that fulfil I_4^4. Due to the symmetry of I_4^4 in x and t, and in y and z, we can assume without loss of generality that $x \leq t$ and $y \leq z$. Therefore, it is sufficient to consider the following 6 cases:

(1) $x \leq y \leq z \leq t$ (2) $x \leq y \leq t \leq z$ (3) $x \leq t \leq y \leq z$

(4) $y \leq x \leq t \leq z$ (5) $y \leq x \leq z \leq t$ (6) $y \leq z \leq x \leq t\,.$

We will restrict ourselves to two of these cases only, the other ones being similar.

The case $x \leq y \leq z \leq t$. If x and t belong to the same summand $\langle a, b, T^* \rangle$, with T^* a t-norm that fulfils I_4^4, then y and z also belong to this summand. Therefore, I_4^4 is equivalent to

$$\begin{aligned}0 \leq x + t &- (a + (b-a)T^*(x', z')) - (a + (b-a)T^*(x', t')) \\ &- (a + (b-a)T^*(y', t')) + (a + (b-a)T^*(y', z')) \leq 1\,,\end{aligned}$$

with $x' = \frac{x-a}{b-a}$, $y' = \frac{y-a}{b-a}$, $z' = \frac{z-a}{b-a}$ and $t' = \frac{t-a}{b-a}$. The latter inequality is equivalent to

$$0 \leq x' + t' - T^*(x', t') - T^*(x', t') - T^*(y', t') + T^*(y', z') \leq \frac{1}{b-a}\,.$$

Since I_4^4 holds for T^* and $1 \leq \frac{1}{b-a}$, the above also holds.

If x and t do not belong to the same summand, then $T(x, t) = x$ and I_4^4 reduces to

$$0 \leq t - T(x, z) - T(y, t) + T(y, z) \leq 1\,. \tag{4.23}$$

It is easy to see that the left inequality is always fulfilled since $t - T(y, t) \geq 0$ and $T(y, z) - T(x, z) \geq 0$. Next, we prove that also the right inequality is fulfilled. Therefore, we split up the proof into different cases.

(i) If y and z do not belong to the same summand (hence $T(y, z) = y$), then also x and z, as well as y and t, do not belong to the same summand (hence $T(x, z) = x$ and $T(y, t) = y$). Therefore, the right part of (4.23) reduces to $t - x \leq 1$, which is obviously fulfilled.

(ii) Suppose y and z belong to the same summand $\langle a, b, T^* \rangle$, with T^* a t-norm that fulfils I_4^4. Again, we have to consider several possibilities:

(a) Also t belongs to this summand, while x does not (hence $T(x,z) = x$). Then the right part of (4.23) is equivalent to

$$t - x - (a + (b-a)T^*(y',t')) + (a + (b-a)T^*(y',z')) \leq 1\,,$$

or also

$$\frac{a-x}{b-a} + t' - T^*(y',t') + T^*(y',z') \leq \frac{1}{b-a}\,.$$

Setting $x = 0$ in I_4^4 and applying it to T^*, we find that

$$0 \leq t' - T^*(y',t') + T^*(y',z') \leq 1\,.$$

It then easily follows that

$$\frac{a-x}{b-a} + t' - T^*(y',t') + T^*(y',z') \leq \frac{a-x}{b-a} + 1 = \frac{b-x}{b-a} \leq \frac{1}{b-a}\,.$$

(b) Also x belongs to this summand, while t does not (hence $T(y,t) = y$). The right part of (4.23) is then equivalent to

$$t - (a + (b-a)T^*(x',z')) - y + (a + (b-a)T^*(y',z')) \leq 1\,,$$

or also

$$\frac{t-a}{b-a} - y' - T^*(x',z') + T^*(y',z') \leq \frac{1}{b-a}\,.$$

Setting $t = 1$ in I_4^4 and applying it to T^*, we find that

$$-1 \leq -y' - T^*(x',z') + T^*(y',z') \leq 0\,.$$

It then follows that

$$\frac{t-a}{b-a} - y' - T^*(x',z') + T^*(y',z') \leq \frac{t-a}{b-a} \leq \frac{1}{b-a}\,.$$

(c) Neither x, nor t belong to this summand (hence $T(x,z) = x$ and $T(y,t) = y$). In this case, the right part of (4.23) is equivalent to

$$t - x - y + (a + (b-a)T^*(y',z')) \leq 1\,,$$

or also

$$\frac{t-x}{b-a} - y' + T^*(y',z') \leq \frac{1}{b-a}\,.$$

Since $-y' + T^*(y',z') \leq 0$, it easily follows that

$$\frac{t-x}{b-a} - y' + T^*(y',z') \leq \frac{t-x}{b-a} \leq \frac{1}{b-a}\,.$$

The case $x \leq y \leq t \leq z$. If x and z belong to the same summand $\langle a, b, T^* \rangle$, with T^* a t-norm that fulfils I_4^4, then y and t also belong to this summand. Therefore, inequality I_4^4 is fulfilled in the same way as in the previous case. If x and z do not belong to the same summand (hence $T(x, z) = x$), then I_4^4 reduces to

$$0 \leq t - T(x, t) - T(y, t) + T(y, z) \leq 1 . \tag{4.24}$$

Again, it is easy to see that the left inequality is always fulfilled since $t - T(x, t) \geq 0$ and $T(y, z) - T(y, t) \geq 0$. Next, we prove that also the right inequality is fulfilled. Therefore, we split up the proof into different cases.

(i) Suppose y and t do not belong to the same summand. The proof is identical to case (i) above.

(ii) Suppose y and t belong to the same summand $\langle a, b, T^* \rangle$, with T^* a t-norm that fulfils I_4^4. Then, we have the following possibilities:

(a) Also z belongs to this summand, while x does not. The proof is identical to case (ii)(a) above.

(b) Also x belongs to this summand, while z does not (hence $T(y, z) = y$). Then the right part of (4.24) is equivalent to

$$t - (a + (b - a)T^*(x', t')) - (a + (b - a)T^*(y', t')) + y \leq 1 ,$$

or also

$$t' - T^*(x', t') - T^*(y', t') + y' \leq \frac{1}{b - a} .$$

Since $T_{\mathbf{L}} \leq T^*$, it holds that $y' + t' - T^*(y', t') \leq 1$ and we can conclude that

$$y' + t' - T^*(x', t') - T^*(y', t') \leq 1 - T^*(x', t') \leq 1 \leq \frac{1}{b - a} .$$

(c) Neither x, nor z belong to this summand (hence $T(x, t) = x$ and $T(y, z) = y$). In this case, the right part of (4.24) is equivalent to

$$t - x - (a + (b - a)T^*(y', t')) + y \leq 1 ,$$

or also

$$y' + t' - T^*(y', t') + \frac{a - x}{b - a} \leq \frac{1}{b - a} .$$

Again, since $T_{\mathbf{L}} \leq T^*$, we can conclude that

$$y' + t' - T^*(y', t') + \frac{a-x}{b-a} \leq 1 + \frac{a-x}{b-a} \leq \frac{b-x}{b-a} \leq \frac{1}{b-a}\,.$$

This completes the proof. □

Moreover, from our experience in proving that ordinal sums preserve the Bell-type inequalities and numerous MATLAB experiments for larger values of n, we can postulate the following more general conjecture.

Conjecture. Consider an inequality of the following form ($n \geq 2$):

$$\sum_{i=1}^{n} a_i\, x_i + \sum_{\substack{i=1 \\ j<i}}^{n} b_{ij}\, T(x_i, x_j) + c \geq 0\,,$$

with $a_i, b_{ij} \in \mathbb{R}$ for all $i = 1, \ldots, n$ and $j < i$. This inequality is preserved under ordinal sums if and only if it is fulfilled by $T_{\mathbf{M}}$, which in turn is equivalent to demanding that $c \geq 0$,

$$a_i + c \geq 0$$

for any i, and

$$\sum_{i=1}^{n} a_i + \sum_{\substack{i=1 \\ j<i}}^{n} b_{ij} + c \geq 0\,.$$

4.8 Bell-type inequalities for parametric t-norm families

In this section, we consider the most important parametric t-norm families and investigate for which values of the parameter involved the corresponding t-norms fulfil a given Bell-type inequality. These families are taken from [52] and are listed in Table 4.3; the subfamilies consisting of copulas are indicated as well. In view of Theorem 4.5, it is sufficient to concentrate on continuous Archimedean t-norms only. As the Mayor–Torrens t-norm family consists of continuous non-Archimedean t-norms, it is excluded from our study, while it does appear in the list of Klement, Mesiar and Pap [52]. Note that all t-norms in Table 4.3 are Archimedean (except for $T_{\mathbf{M}}$, which appears as a limit case in some families).

In the following subsections, we consider the Bell-type inequalities one by one and identify for each of the families in Table 4.3, the range of parameters for which the corresponding t-norms fulfil the given inequality. The results of this study are summarized in Table 4.4. The delimiting parameter values for the Frank t-norm family are taken from Section 4.6.

4.8.1 Inequalities I_2^1, I_3^2, I_4^4 and I_4^5

Thanks to Theorems 4.2 and 4.4, we already know that inequalities I_2^1, I_3^2, I_4^4 and I_4^5 are fulfilled for any commutative copula. We have verified that for the parametric families considered, none of its non-copula members satisfies any of these inequalities.

It can easily be shown that inside the unit square $]0,1[^2$ the first-order derivatives of the function $f(x,y) = x + y - T(x,y) - 1$, with T belonging to one of the families in Table 4.3, can only be zero in the symmetric case $x = y$. Therefore, inequality I_2^1 is equivalent to

$$2x - T(x,x) \leq 1\,. \tag{4.25}$$

Let us consider for instance the Yager family and the Schweizer–Sklar family. The choice of a t-norm family is chosen at random and such that all t-norm families are given a chance in the following subsections.

Yager t-norm family. Considering a t-norm belonging to the Yager family, the above inequality can be written explicitly as

$$2x - \max(0, 1 - 2^{1/\lambda}(1-x)) - 1 \leq 0\,. \tag{4.26}$$

Obviously, we have to consider two cases:

(i) The case $\max(0, 1-2^{1/\lambda}(1-x)) = 0$. It then holds that $1-2^{1/\lambda}(1-x) \leq 0$ and the latter inequality reads $2x - 1 \leq 0$, which holds for $\lambda \geq 1$.

(ii) The case $\max(0, 1-2^{1/\lambda}(1-x)) > 0$. In that case, inequality (4.26) reads $(2 - 2^{1/\lambda})(x-1) \leq 0$. It is easy to see that this inequality is fulfilled if $2 - 2^{1/\lambda} \geq 0$, or equivalently, if $\lambda \geq 1$.

Both cases lead to the same restriction on λ, and we can conclude that inside the Yager family, inequality I_2^1 only holds for its copula members.

T-norm family	$T_\lambda(x,y) =$	Copulas for
Frank ($\lambda \in [0,+\infty]$)	$\begin{cases} T_{\mathbf{M}}(x,y) & \text{, if } \lambda = 0 \\ T_{\mathbf{P}}(x,y) & \text{, if } \lambda = 1 \\ T_{\mathbf{L}}(x,y) & \text{, if } \lambda = \infty \\ \log_\lambda(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & \text{, otherwise} \end{cases}$	$\lambda \in [0,+\infty]$
Hamacher ($\lambda \in [0,+\infty]$)	$\begin{cases} T_{\mathbf{D}}(x,y) & \text{, if } \lambda = \infty \\ 0 & \text{, if } \lambda = x = y = 0 \\ \frac{xy}{\lambda+(1-\lambda)(x+y-xy)} & \text{, otherwise} \end{cases}$	$\lambda \in [0,2]$
Schweizer–Sklar ($\lambda \in [-\infty,+\infty]$)	$\begin{cases} T_{\mathbf{M}}(x,y) & \text{, if } \lambda = -\infty \\ T_{\mathbf{P}}(x,y) & \text{, if } \lambda = 0 \\ T_{\mathbf{D}}(x,y) & \text{, if } \lambda = +\infty \\ (\max(x^\lambda + y^\lambda - 1, 0))^{\frac{1}{\lambda}} & \text{, otherwise} \end{cases}$	$\lambda \in [-\infty,1]$

Table 4.3: Different t-norm families used throughout this work.

Sugeno–Weber ($\lambda \in [-1, +\infty]$)	$\begin{cases} T_{\mathbf{D}}(x,y) & \text{, if } \lambda = -1 \\ T_{\mathbf{P}}(x,y) & \text{, if } \lambda = +\infty \\ \max(\frac{x+y-1+\lambda xy}{1+\lambda}, 0) & \text{, otherwise} \end{cases}$	$\lambda \in [0, +\infty]$
Dombi ($\lambda \in [0, +\infty]$)	$\begin{cases} T_{\mathbf{D}}(x,y) & \text{, if } \lambda = 0 \\ T_{\mathbf{M}}(x,y) & \text{, if } \lambda = +\infty \\ \frac{1}{1+((\frac{1-x}{x})^\lambda+(\frac{1-y}{y})^\lambda)^{\frac{1}{\lambda}}} & \text{, otherwise} \end{cases}$	$\lambda \in [1, +\infty]$
Aczel–Alsina ($\lambda \in [0, +\infty]$)	$\begin{cases} T_{\mathbf{D}}(x,y) & \text{, if } \lambda = 0 \\ T_{\mathbf{M}}(x,y) & \text{, if } \lambda = +\infty \\ \exp^{-((-\log x)^\lambda+(-\log y)^\lambda)^{\frac{1}{\lambda}}} & \text{, otherwise} \end{cases}$	$\lambda \in [1, +\infty]$
Yager ($\lambda \in [0, +\infty]$)	$\begin{cases} T_{\mathbf{D}}(x,y) & \text{, if } \lambda = 0 \\ T_{\mathbf{M}}(x,y) & \text{, if } \lambda = +\infty \\ \max(1-((1-x)^\lambda+(1-y)^\lambda)^{\frac{1}{\lambda}}, 0) & \text{, otherwise} \end{cases}$	$\lambda \in [1, +\infty]$

Table 4.3: Different t-norm families used throughout this work (continued).

Schweizer–Sklar t-norm family. Considering a t-norm belonging to the Schweizer–Sklar family, the above inequality can be written explicitly as

$$2x - (\max(0, 2x^\lambda - 1))^{\frac{1}{\lambda}} - 1 \leq 0\,. \tag{4.27}$$

Again, we have to consider two cases:

(i) The case $\max(0, 2x^\lambda - 1) = 0$. It then holds that $2x^\lambda - 1 \leq 0$ and the latter inequality reads $2x - 1 \leq 0$, which holds for $\lambda \leq 1$.

(ii) The case $\max(0, 2x^\lambda - 1) > 0$. In that case, inequality (4.27) reads $2x - 1 - (2x^\lambda - 1)^{\frac{1}{\lambda}} \leq 0$. The left-hand side of this inequality defines a function f, with

$$f(x) = 2x - 1 - (2x^\lambda - 1)^{\frac{1}{\lambda}}\,.$$

The first-order derivative of f equals

$$f'(x) = 2 - 2\frac{x^\lambda(2x^\lambda - 1)^{\frac{1}{\lambda}}}{x(2x^\lambda - 1)}\,.$$

Solving $f'(x) = 0$, we find that f reaches an extremal value in $x_{s_1} = 0$, $x_{s_2} = 1$ or $x_{s_3} = \left(\frac{1}{2}\right)^{\frac{1}{\lambda}}$. It is easy to see that $f(x_{s_1}) \leq 0$ and $f(x_{s_2}) \leq 0$ for any λ. Then,

$$\begin{aligned} f(x_{s_3}) \leq 0 &\Leftrightarrow 2\left(\frac{1}{2}\right)^{\frac{1}{\lambda}} - 1 \leq 0 \\ &\Leftrightarrow \left(\frac{1}{2}\right)^{\frac{1}{\lambda}} \leq \frac{1}{2} \\ &\Leftrightarrow \lambda \leq 1\,. \end{aligned}$$

Both cases lead to the same restriction on λ, and we can conclude that inside the Schweizer–Sklar family, inequality I_2^1 only holds for its copula members.

Similarly, inequality I_4^4 is equivalent to

$$2x - 2T(x, y) - T(x, x) + T(y, y) \leq 1\,,$$

while inequality I_4^5 is equivalent to

$$-2x + 4T(x, y) - T(x, x) + T(y, y) \leq 0\,.$$

Such a simplified equivalent inequality does not exist for inequality I_3^2. The verification for inequalities I_3^2, I_4^4 and I_4^5 was done in a numerical way.

4.8.2 Inequality I_3^3

It can easily be shown that inside the unit cube $]0,1[^3$ the first-order derivatives of the function

$$f(x,y,z) = x + y + z - T(x,y) - T(x,z) - T(y,z) - 1\,,$$

with T belonging to one of the families in Table 4.3, can only be zero in the symmetric case $x = y = z$. Therefore, inequality I_3^3 is equivalent to

$$3x - 3T(x,x) \leq 1\,. \tag{4.28}$$

We focus our attention on the Dombi and Aczel–Alsina t-norm families. The delimiting parameter values for the other families can be obtained in a similar way. In case no exact solution to a given problem was found, the help of Maple was called in to find a numerical solution.

Dombi t-norm family. For the Dombi t-norm family, inequality (4.28) reads explicitly

$$1 - 3x + \frac{3x}{x + 2^{1/\lambda}(1-x)} \geq 0\,.$$

Reducing the left-hand side of this inequality to the same (positive) denominator, it is sufficient to study the numerator, which defines a quadratic function g:

$$g(x) = 3x^2(2^{1/\lambda} - 1) - 4x(2^{1/\lambda} - 1) + 2^{1/\lambda}\,.$$

We determine the values of λ such that $g(x) \geq 0$ for any $x \in [0,1]$. Solving $g'(x) = 0$, we find that g reaches an extremal value in $x_s = 2/3$. Moreover, it is easy to see that the discriminant of g (i.e. $2 \cdot 2^{1/\lambda} - 10 \cdot 2^{1/\lambda} + 8$) is negative or zero when $\lambda \geq 1/2$ and in this case $g(x) \geq 0$ for any $x \in [0,1]$. On the other hand, the discriminant of g is positive when $\lambda < 1/2$. In this case $g(2/3) < 0$ and we can conclude that the sign of $g(x)$ will change in the interval $[0,1]$. Therefore, inequality I_3^3 holds for any $\lambda \in [1/2, +\infty[$.

Aczel–Alsina t-norm family. For the Aczel–Alsina t-norm family, inequality (4.28) reads explicitly

$$3x - 3x^{(2^{\frac{1}{\lambda}})} - 1 \leq 0\,.$$

Substituting $2^{\frac{1}{\lambda}}$ by t and dividing the left-hand side of the latter inequality by 3 defines a function g:

$$g(x) = x - x^t - \frac{1}{3}\,.$$

We determine the values of t such that $g(x) \geq 0$ for any $x \in [0,1]$. Solving $g'(x) = 0$, we find that g reaches an extremal value in $x_s = (\frac{1}{t})^{\frac{1}{t-1}}$. Substituting this value for x_s in g, we obtain a function h

$$h(t) = \left(\frac{1}{t}\right)^{\frac{1}{t-1}} - \left(\frac{1}{t}\right)^{\frac{t}{t-1}} - \frac{1}{3}\,.$$

Solving $h(t) = 0$ numerically with the help of Maple, we obtain $t = 2.5581$ and so $\lambda = 0.7379$. Therefore, we can conclude that inequality (4.28) is fulfilled when $\lambda \geq 0.7379$.

Note that for the Dombi and Aczel–Alsina families inequality I_3^3 is fulfilled for all of its copula members. This is for instance not the case for the Frank family. Indeed, although all Frank t-norms are copulas, inequality I_3^3 is only fulfilled for $\lambda \in [0, 9 + 4\sqrt{5}]$.

4.8.3 Inequalities I_4^6 and I_4^7

In this subsection, we consider inequalities I_4^6 and I_4^7. Again, inside the unit hypercube $]0,1[^4$ the first-order derivatives of the function

$$\begin{aligned} f(x,y,z,t) = {} & x + y + z + t - T(x,y) - T(x,z) - T(x,t) \\ & - T(y,z) - T(y,t) - T(z,t)\,, \end{aligned}$$

with T belonging to one of the families in Table 4.3, can only be zero in the symmetric case $x = y = z = t$. Therefore, inequality I_4^6 is equivalent to

$$4x - 6T(x,x) \leq 1\,. \tag{4.29}$$

Let us consider for instance the Hamacher and Yager t-norm families.

Hamacher t-norm family. For the Hamacher family, the above inequality then reads explicitly:

$$4x - 6\frac{x^2}{\lambda + (1-\lambda)(2x - x^2)} - 1 \leq 0\,.$$

Reducing the left-hand side of this inequality to the same (positive) denominator, it is sufficient to study the numerator, which defines a cubic function g:

$$g(x) = 4(\lambda - 1)x^3 - 3(3\lambda - 1)x^2 + 2(3\lambda - 1)x - \lambda\,.$$

We determine the values of λ such that $g(x) \leq 0$ for any $x \in [0,1]$. For $\lambda = 1$, the function g reduces to the quadratic function h, given by

$$h(x) = -6x^2 + 4x - 1\,.$$

Since its discriminant is negative, the function h is negative for any $x \in [0,1]$.

Now consider $\lambda \neq 1$. The first-order derivative of g is given by

$$g'(x) = 12(\lambda-1)x^2 - 6(3\lambda-1)x + 2(3\lambda-1)\,.$$

The discriminant of this quadratic function, i.e. $3(3\lambda-1)(\lambda+5)$, is positive or equal to zero when $\lambda \geq 1/3$. In that case, the function g' has two real roots. We consider two different cases:

(i) The case $1/3 \leq \lambda < 1$: the smallest root of g' is always smaller than 0, while the other one, say x_s, belongs to the interval $[0,1]$. Therefore, it is necessary and sufficient that $g(x_s) \leq 0$ to guarantee that $g(x) \leq 0$ for any $x \in [0,1]$. Invoking Maple, we can conclude that $g(x_s) \leq 0$.

(ii) The case $1 < \lambda$: the smallest root of g', say x_s, belongs to the interval $[0,1]$, while the other one is always greater than 1. Again, it is necessary and sufficient that $g(x_s) \leq 0$ to guarantee that $g(x) \leq 0$ for any $x \in [0,1]$. Invoking Maple, we can conclude that $g(x_s) \leq 0$ when $\lambda \leq 2.6529$.

Therefore, we can conclude that inequality I_4^6 is fulfilled for $\lambda \leq 2.6529$.

Similarly, inequality I_4^7 is equivalent to

$$8x - 6T(x,x) \leq 3\,, \tag{4.30}$$

which is in the Hamacher family fulfilled when $\lambda \leq 2.222$.

Yager t-norm family. Considering a t-norm belonging to the Yager family, inequality (4.29) can be written explicitly as

$$4x - 6\max(0, 1 - 2^{\frac{1}{\lambda}}(1-x)) - 1 \leq 0\,. \tag{4.31}$$

Obviously, we have to consider two cases:

(i) The case $\max(0, 1-2^{1/\lambda}(1-x)) = 0$. It then holds that $1-2^{1/\lambda}(1-x) \leq 0$ and the latter inequality reads $4x - 1 \leq 0$, which holds for $\lambda \geq \frac{\ln 2}{\ln 4 - \ln 3}$.

(ii) The case $\max(0, 1 - 2^{1/\lambda}(1-x)) > 0$. In that case, inequality (4.31) reads $x(4 - 6 \cdot 2^{\frac{1}{\lambda}}) + 6 \cdot 2^{\frac{1}{\lambda}} - 7 \leq 0$. The left-hand side of this inequality defines a linear function f, with

$$f(x) = x(4 - 6 \cdot 2^{\frac{1}{\lambda}}) + 6 \cdot 2^{\frac{1}{\lambda}} - 7\,.$$

Since $4 - 6 \cdot 2^{\frac{1}{\lambda}} \leq 0$ for any $\lambda \leq 0$, it's easy to verify that $f(x) \geq 0$ if $\lambda \geq \frac{\ln 2}{\ln 7 - \ln 6}$.

We can conclude that inside the Yager family, inequality (4.29) holds for $\lambda \geq \frac{\ln 2}{\ln 7 - \ln 6}$.

In the same way, we obtain $\lambda \geq \frac{\ln 2}{\ln 3 - \ln 2}$ such that inequality (4.30) is satisfied.

Note that for the Hamacher family inequalities I_4^6 and I_4^7 are fulfilled for all of its copula members, while this is clearly not the case for the Yager or the Frank family.

4.8.4 Inequalities I_4^8 and I_4^9

Finally, we consider inequalities I_4^8 and I_4^9. For inequality I_4^8, for instance, the first-order derivatives of the function

$$f(x,y,z,t) = -x + T(x,y) + T(x,z) + T(x,t) - T(y,z) - T(y,t) - T(z,t)\,,$$

with T belonging to one of the families in Table 4.3, can only be zero in the symmetric case $y = z = t$. This renders inequality I_4^8 equivalent to

$$-x + 3T(x,y) - 3T(y,y) \leq 0\,.$$

This time, we consider the Sugeno–Weber family. The above inequality then reads:

$$-x + 3\,\max\left(0, \frac{x + y - 1 + \lambda xy}{1+\lambda}\right) - 3\,\max\left(0, \frac{2y - 1 + \lambda y^2}{1+\lambda}\right) \leq 0\,. \tag{4.32}$$

We distinguish four different cases. When both maxima are equal to zero, inequality (4.32) reduces to the trivial inequality $-x \leq 0$. Also, when $\max(0, \frac{x+y-1+\lambda xy}{1+\lambda}) = 0$, inequality (4.32) reduces to $-x - 3\left(\frac{2y-1+\lambda y^2}{1+\lambda}\right) \leq 0$ which is easily verified for any $\lambda > -1$. The third case being similar to the previous one, it only remains to consider the

Family	$I_2^1, I_3^2, I_4^4, I_4^5$	I_3^3, I_4^8, I_4^9	I_4^6	I_4^7
Frank	$[0, +\infty]$	$[0, 9 + 4\sqrt{5}]$	$[0, 9.2946]$	$[0, 9.2946]$
Hamacher	$[0, 2]$	$[0, 2.9386]$	$[0, 2.6529]$	$[0, 2.2220]$
Schweizer-Sklar	$[-\infty, 1]$	$[-\infty, \frac{1}{2}]$	$[-\infty, 0.3435]$	$[-\infty, \frac{1}{2}]$
Sugeno-Weber	$[0, +\infty]$	$[3, +\infty]$	$[8, +\infty]$	$[2, +\infty]$
Dombi	$[1, +\infty]$	$[\frac{1}{2}, +\infty]$	$[\frac{\ln 2}{\ln(3+\frac{4}{3}\sqrt{2})}, +\infty]$	$[\frac{\ln 2}{\ln 3}, +\infty]$
Aczel-Alsina	$[1, +\infty]$	$[0.7379, +\infty]$	$[0.7533, +\infty]$	$[0.8201, +\infty]$
Yager	$[1, +\infty]$	$[\frac{\ln 2}{2\ln 2 - \ln 3}, +\infty]$	$[\frac{\ln 2}{\ln 7 - \ln 6}, +\infty]$	$[\frac{\ln 2}{\ln 3 - \ln 2}, +\infty]$

Table 4.4: Conditions on the parameter λ.

case that both maxima are different from zero. In that case, inequality (4.32) reads:

$$-x + 3\frac{x + y - 1 + \lambda xy}{1 + \lambda} - 3\frac{2y - 1 + \lambda y^2}{1 + \lambda} \le 0\,. \tag{4.33}$$

Note that this inequality only needs to be considered in the domain enclosed by the boundaries $x = 1$, $y = 1$, $x + y - 1 + \lambda\, xy = 0$ and $y = \frac{-1+\sqrt{1+\lambda}}{\lambda}$. If we reduce the left-hand side of this inequality to the same (positive) denominator, the numerator determines a two-place function g:

$$g(x, y) = -3\lambda y^2 + 3\lambda xy - 3y + (2 - \lambda)x\,.$$

We determine the values of λ such that $g(x, y) \le 0$ for any $(x, y) \in [0, 1]^2$. In order to find the stationary points of g, we set the first-order derivatives of g equal to zero, and obtain:

$$\begin{aligned} g_x(x, y) &= 3\lambda y + 2 - \lambda = 0\,,\\ g_y(x, y) &= -6\lambda y + 3\lambda x - 3 = 0\,. \end{aligned}$$

Solving this system of equations we obtain a single solution

$$(x_s, y_s) = \left(\frac{2\lambda - 1}{3\lambda}, \frac{\lambda - 2}{3\lambda}\right)\,,$$

which is a stationary point of g. Furthermore, it holds that

$$g(x_s, y_s) = \frac{2 - \lambda}{3\lambda}\,.$$

We need to verify whether this stationary point is a minimum, maximum or saddle point. To that end, we compute the second-order derivatives of g:

$$\begin{aligned} g_{xx}(x_s, y_s) &= 0\,,\\ g_{yy}(x_s, y_s) &= -6\lambda\,,\\ g_{xy}(x_s, y_s) &= 3\lambda\,, \end{aligned}$$

from which it follows that the determinant of these derivatives of g in the stationary point (x_s, y_s) is given by

$$A_2 = [g_{xx}g_{yy} - g_{xy}^2](x_s, y_s) = -9\lambda^2\,.$$

Since $A_2 < 0$ for any $\lambda \in]-1, +\infty[$, the stationary point (x_s, y_s) is neither a minimum, nor a maximum (it is a saddle point). Therefore, the maximum of g will be reached on the boundaries of the domain of g.

(i) On the boundary $y = \frac{-1+\sqrt{1+\lambda}}{\lambda}$, we obtain a linear function h in x:

$$h(x) = x\left(\frac{3\sqrt{1+\lambda}-1-\lambda}{1+\lambda}\right) + 3\frac{\lambda\sqrt{1+\lambda}-1-\lambda}{\lambda(1+\lambda)}\,.$$

We determine the values of λ such that $h(x) \leq 0$ for any $x \in [\frac{1+\lambda-\sqrt{1+\lambda}}{\lambda\sqrt{1+\lambda}}, 1]$. It is easily verified that h is an increasing function when $-1 < \lambda \leq 8$, while h is decreasing when $\lambda > 8$. If h is increasing, $h(1)$ should be negative in order that $h(x) \leq 0$, for any $x \in [\frac{1+\lambda-\sqrt{1+\lambda}}{\lambda\sqrt{1+\lambda}}, 1]$. It is easy to see that $h(1) \leq 0$ when $3 \leq \lambda \leq 8$. In the same way, if h is decreasing, $h(\frac{1+\lambda-\sqrt{1+\lambda}}{\lambda\sqrt{1+\lambda}})$ should be negative. This is the case when $\lambda > 8$. Therefore, we can conclude that $h(x) \leq 0$ for any $x \in [\frac{1+\lambda-\sqrt{1+\lambda}}{\lambda\sqrt{1+\lambda}}, 1]$ when $\lambda \geq 3$.

(ii) Similarly, on the boundary $x + y - 1 + \lambda xy = 0$ (or equivalently, $y = \frac{x-1}{-\lambda x-1}$), we obtain another function h in x:

$$h(x) = \frac{-\lambda^2(1+\lambda)x^3 + \lambda(1+\lambda)x^2 + 5(1+\lambda)x - 3(\lambda+1)}{(\lambda x+1)^2}\,.$$

In the same way, it holds that $h(x) \leq 0$ for any $x \in [0, \frac{1+\lambda-\sqrt{1+\lambda}}{\lambda\sqrt{1+\lambda}}]$ when $\lambda \geq 175/81$.

(iii) It is easy to see that on the boundary $y = 1$, inequality (4.33) reduces to $(1+\lambda)(2x-3) \leq 0$ and is always fulfilled, while the boundary $x = 1$ leads to inequality (4.28), which is certainly fulfilled when $\lambda \geq 3$.

Summarizing all cases above, we can conclude that inequality I_4^8 is satisfied when $\lambda \geq 3$.

Similarly, inequality I_4^9 is equivalent to

$$3x - 2y - 3T(x,x) + 3T(x,y) \leq 1\,,$$

which is in the Sugeno–Weber family fulfilled when $\lambda \geq 3$.

Note that it is not that easy to find the delimiting parameter values for all families. In most cases, the resulting functions were too complicated to find analytical solutions or even numerical ones. As a way out, we used contour plots to conclude that no extrema occurred inside the unit square $]0,1[^2$, and in some cases, only saddle points (see also Intermezzo 4.2). Hence, for all families an extremum will be reached on the

Intermezzo 4.2 Critical points of $f(x, y)$.

A function $z = f(x, y)$ defines a three-dimensional surface. Like its analog in two dimensions, $f(x, y)$ may have maximum and minimum values. The critical points (minima, maxima or saddle points) can be found by taking the partial derivatives of f and solving the following system for x and y:

$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = 0 \\ \frac{\partial f}{\partial y}(x, y) = 0 \end{cases}.$$

The (analytical) solution to this system of equations may be difficult to find, even when using numerical tools, like Maple. Therefore, we are thrown back on other methods. Although, there exist various numerical optimization techniques to solve this problem, also contour lines can be used to gain more insight in the behavior of f. Therefore, we represent the "landscape" of the surface $z = f(x, y)$ by contour lines, which are curves in the $(x, y)-$plane on which $f(x, y)$ takes different constant values.

Around a maximum, the value of $f(x, y)$ is always smaller than its value z^* at the maximum. The contours are closed loops around the stationary point (see Figure 4.4). Around a minimum, $f(x, y) > z^*$ and again the contours are closed loops around the stationary point (see Figure 4.5). The representation of a saddle point by contour lines has the characteristic appearance as depicted in Figure 4.6.

boundaries of the unit square $[0, 1]^2$ (which can be seen from Table 4.5). Therefore, the parameter values such that I_4^8 and I_4^9 are satisfied, are the same as the ones obtained for I_3^3. These results are summarized in Table 4.4.

Example 4.5 Consider the Hamacher t-norm family. Inequality I_4^8 reads as

$$-x + \frac{3xy}{\lambda + (1-\lambda)(x + y - xy)} - \frac{3y^2}{\lambda + (1-\lambda)(2y - y^2)} \leq 0\,.$$

For $\lambda < 1$, we obtain contour plots as in Figure 4.7, while for $\lambda \geq 1$ contour plots as in Figure 4.8 are obtained. In both cases, it is easy to see that an extremum will be reached on the boundary of the unit interval.

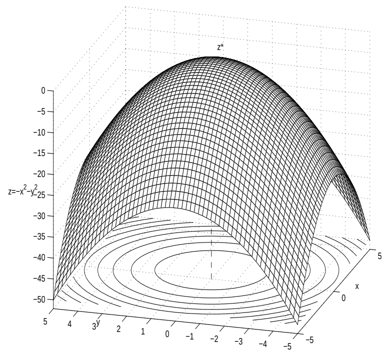


Figure 4.4: The representation of a maximum by contour lines.

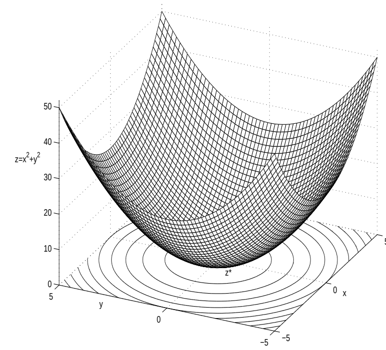


Figure 4.5: The representation of a minimum by contour lines.

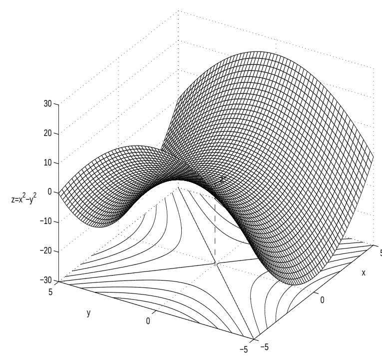


Figure 4.6: The representation of a saddle point by contour lines.

4.9 A family of Bell-type inequalities

Taking a closer look at the inequalities of type $c_1\,x - c_2\,T(x,x) \leq c_3$, with constants $c_1, c_2, c_3 \geq 0$, such as the inequality $3x - 3T(x,x) \leq 1$, suggests the following general form, $n \geq 2$:

$$nx - \binom{n}{2}\,T(x,x) \leq 1\,. \tag{4.34}$$

For $n = 2$, we obtain the inequality $2x - T(x,x) \leq 1$ and for $n = 3$, we retrieve the inequality $3x - 3T(x,x) \leq 1$, i.e. the necessary and sufficient condition for inequality I_3^3 to hold for a Frank t-norm. Similarly, for $n = 4$, we find $4x - 6T(x,x) \leq 1$, i.e. the equivalent of inequality I_4^6. We can prove the following theorem:

Boundaries	I_4^8	I_4^9
$x = 0$	$-2y \leq 1$	$-3T(y, y) \leq 0$
$x = 1$	$y \leq 1$	$3y - 3T(y, y) \leq 1$
$y = 0$	$3x - 3T(x, x) \leq 1$	$-x \leq 0$
$y = 1$	$3x - 3T(x, x) \leq 0$	$2x - 3 \leq 0$

Table 4.5: Inequalities obtained on the boundaries of the unit square for I_4^8 and I_4^9.

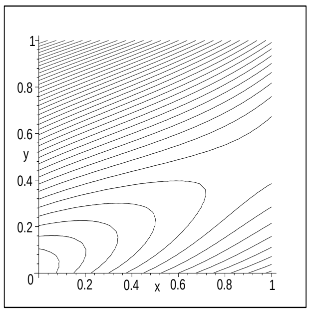


Figure 4.7: Contour plot for a Hamacher t-norm with $\lambda = 0.2 < 1$.

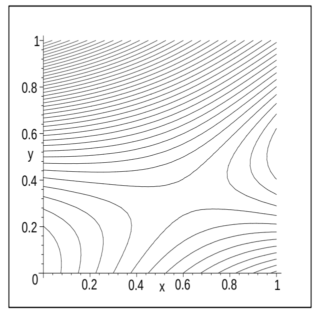


Figure 4.8: Contour plot for a Hamacher t-norm with $\lambda = 2 \geq 1$.

Theorem 4.6 *The only Frank t-norms for which inequality (4.34) is fulfilled for all $n \geq 2$ are the t-norms $T_\lambda^{\mathbf{F}}$ between the algebraic product $T_{\mathbf{P}}$ and the minimum operator $T_{\mathbf{M}}$ (i.e. with $\lambda \in [0, 1]$).*

Proof. Let us first consider the case $T = T_{\mathbf{P}}$, i.e. $\lambda = 1$. The expression $nx - n(n-1)x^2/2 - 1$ reaches a maximum in $x_s = 1/(n-1)$, a point located in $[0, 1]$ for all $n \geq 2$, and the maximum value being $-(n-2)/2(n-1)$, inequality (4.34) is clearly fulfilled for all $n \geq 2$. For all $T \geq T_{\mathbf{P}}$, in particular for all members of the Frank family with $\lambda \in [0, 1[$, inequality (4.34) is then obviously also fulfilled.

We still have to investigate the case $\lambda > 1$. We have to determine the values $\lambda > 1$ for which

$$nx - \frac{n(n-1)}{2} \log_\lambda \left(1 + \frac{(\lambda^x - 1)^2}{\lambda - 1}\right) \leq 1\,,$$

or, equivalently, introducing the notation $s = \lambda^x$, the values $\lambda > 1$ for which

$$s^n \left(1 + \frac{(s-1)^2}{\lambda - 1}\right)^{-n(n-1)/2} \leq \lambda\,,$$

or

$$s^{-2/(n-1)}\left(1+\frac{(s-1)^2}{\lambda-1}\right)\le \lambda^{-2/(n(n-1))}\,.$$

In the limit of $n\to\infty$, this condition becomes

$$1+\left(1+\frac{(s-1)^2}{\lambda-1}\right)\le 1\,,$$

which is clearly not fulfilled for all $s\in[1,\lambda]$, and hence not for all $x\in[0,1]$. $\square$

When generating all Bell-type inequalities for any $n>0$, we will obtain inequalities of the form (4.34) as necessary and sufficient conditions for the Frank t-norm family. From this, we can conclude that within this family at most all t-norms between the minimum operator $T_{\mathbf{M}}$ and the algebraic product $T_{\mathbf{P}}$ will satisfy all Bell-type inequalities.

In general, the algebraic product $T_{\mathbf{P}}$ is not the smallest t-norm that satisfies inequalities (4.34). This is confirmed by the following example.

Example 4.6 The Hamacher t-norm with $\lambda = 2$, i.e. $T_2^{\mathbf{H}}(x,y) = \frac{xy}{2-x-y+xy}$, which is smaller than the algebraic product ($T_2^{\mathbf{H}} < T_1^{\mathbf{H}} = T_{\mathbf{P}}$), fulfills inequality (4.34) for any $n\ge 2$.

Proof. Writing inequality (4.34) explicitly, we obtain

$$nx-\frac{n(n-1)}{2}\frac{x^2}{x^2-2x+2}-1\le 0\,.$$

If we reduce the left-hand side of this inequality to the same (positive) denominator, then the numerator determines a function f:

$$f(x)=2nx^3-(n^2+3n+2)x^2+4(n+1)x-4\,.$$

For $n=2$, this function reduces to $f(x)=4(x-1)^3$ and obviously $f(x)\le 0$ for any $x\in[0,1]$. Now suppose $n\ge 3$. The first-order derivative of f is given by

$$f'(x)=6nx^2-(2n^2+6n+4)x+4(n+1)\,.$$

Next we solve the equation $f'(x)=0$. Since $n\ge 3$, the discriminant of this quadratic function ($D=(n-2)(n+1)(n^2+7n-2)$) is always positive, and therefore f' has two real roots. It is easy to see that one root of this equation, say x_s, lies between 0 and 1, while the second one is always greater than 1. Therefore, a necessary and sufficient condition in order that $f(x) \leq 0$ for any $x \in [0,1]$ is that $f(x_s)$ should be negative for any $n \geq 3$. Straightforward computation yields

$$f(x_s) = -\frac{(n-2)}{54n^2}\Big((n+1)(n^2+7n-2)\sqrt{D} \\ + \ (n-1)(n^4+12n^3+31n^2-12n+4) \Big) \ ,$$

with $D = n^4 + 6n^3 - 11n^2 - 12n + 4$, which is negative for all $n \geq 3$. This completes our proof. □

4.10 Conclusions and indications for future research

In this chapter, we have described all Bell-type inequalities concerning four events in which at most two events are intersected at the same time. We have rewritten them in the context of fuzzy probability calculus, or equivalently, to the context of basic scalar cardinalities and have proven that some inequalities are fulfilled for (quasi-)copulas. Moreover, considering the Frank t-norm family, for each of the Bell-type inequalities we have identified all parameter values such that it is fulfilled. These results are summarized in Table 4.2.

We have also studied in detail the Bell-type inequalities for continuous t-norms. A major contribution of this chapter is that ordinal sums preserves the Bell-type inequalities, which was the motivation for studying continuous Archimedean t-norms only. As general results based on additive generators are unlikely to be obtained, we have discussed in an exhaustive way the major parametric t-norm families. The results of this study are recapitulated in Table 4.4. Finally, for a particular form of these inequalities, we have shown that the algebraic product $T_{\mathbf{P}}$ is not the smallest t-norm fulfilling them.

In this chapter, we have rewritten the Bell inequalities in the context of basic scalar cardinalities. But, the basic scalar cardinality is only one of the possibilities to define the cardinality of a fuzzy set. It is interesting to ask whether the Bell inequalities would also be fulfilled for integer cardinalities or even fuzzy cardinalities.

Chapter 5

Inequalities on scalar cardinalities

5.1 Introduction

In Chapter 3, identities on cardinalities were discussed in detail. However, not only identities on cardinalities are of interest, but even more, inequalities. The basic inequalities that come to mind are the Bell inequalities. As the Bell inequalities deal with classical probabilities, it is straightforward to formulate analogous inequalities on cardinalities. We have already investigated in depth the validity of these Bell inequalities on cardinalities in a fuzzy context, mainly focusing on the choice of logical connectives involved. The purpose of this chapter is twofold. On the one hand we will demonstrate that the Bell inequalities are of particular interest in the context of cardinalities of fuzzy sets. On the other hand, we want to show how the results on the fuzzified Bell inequalities can be exploited to develop a framework in which the validity of more general inequalities on fuzzy cardinalities can be checked easily, thus avoiding unnecessary repetitions of tedious calculations.

5.2 Bell-type inequalities for cardinalities

As Bell inequalities deal with classical probabilities, it is straightforward to formulate analogous inequalities on cardinalities. Consider a finite universe X of cardinality n, then the classical probability $P(A)$ is given by $|A|/n$ and the Bell inequalities can be rewritten in the following form:

$$
\begin{aligned}
B_2^1 : & \quad |A| + |B| - n \leq |A \cap B| \leq \min(|A|, |B|)\,, \\
B_3^2 : & \quad 0 \leq |A| - |A \cap B| - |A \cap C| + |B \cap C|\,, \\
B_3^3 : & \; |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| \leq n\,,
\end{aligned}
$$

for any $A, B, C \in \mathcal{P}(X)$. Note that the vertices of the polytope now correspond to the extreme cases $A, B, C \in \{\emptyset, X\}$.

Remark 5.1 *Remark that the double inequality B_2^1 follows from inequality B_3^2 by putting $A = X$, resp. $C = \emptyset$.*

Although it is possible to rewrite all Bell inequalities in terms of cardinalities, we don't mention them explicitly since only inequalities B_2^1–B_3^3 will be of further interest in this work.

The Bell-type inequalities B_2^1–B_3^3 are of particular interest in the context of cardinalities of fuzzy sets. Consider for instance the classical inequalities

$$|A \setminus C| \leq |A \setminus B| + |B \setminus C|\,, \tag{5.1}$$

$$|A \bigtriangleup C| \leq |A \bigtriangleup B| + |B \bigtriangleup C|\,. \tag{5.2}$$

They can be rewritten less elegantly in terms of intersections as follows

$$
\begin{aligned}
& 0 \leq |B| - |A \cap B| - |B \cap C| + |A \cap C|\,, \\
& 0 \leq 2(|B| - |A \cap B| - |B \cap C| + |A \cap C|)\,,
\end{aligned}
$$

respectively. Modelling fuzzy set intersection by means of a commutative conjunctor I, the latter inequalities will remain valid for fuzzy sets if and only if I satisfies

$$0 \leq x - I(x, y) - I(x, z) + I(y, z)\,,$$

i.e. if and only if I satisfies inequality I_3^2, for instance, when using a commutative quasi-copula I.

In a setting where additionally $|A \setminus B| = |A| - |A \cap B|$ and $|A \bigtriangleup B| = |A| + |B| - 2|A \cap B|$ hold for cardinalities of fuzzy sets, we can also conclude that inequalities (5.1) and (5.2) will hold for fuzzy cardinalities if inequality I_3^2 is satisfied. These observations have proven to be crucial for the design of fuzzification schemes for cardinality-based similarity measures which preserve transitivity. In [19], the authors propose a class of fuzzification schemes, whereby the intersection of two

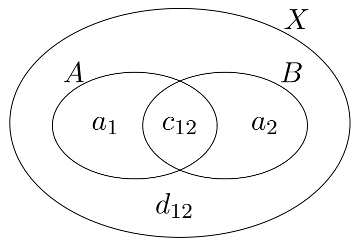


Figure 5.1: Notations for cardinalities associated with two ordinary subsets A and B of a finite universe X.

fuzzy sets is modelled by a Frank t-norm, that can be used to translate cardinality-based similarity measures for ordinary sets into fuzzy similarity measures preserving transitivity.

Similarly as for inequality (5.2), we can restate the classical inequality

$$|A_1 \triangle A_n| \leq \sum_{i=1}^{n-1} |A_i \triangle A_{i+1}|$$

which is valid for ordinary sets $A_1, \ldots, A_n$ in a finite universe X of cardinality n, in terms of intersections only. The left-hand side of (4.9) can then be seen as a necessary and sufficient condition for this new inequality to remain true for fuzzy sets.

5.3 Meta-theorems

In this section we present two meta-theorems, the first one involving two fuzzy sets and their intersection, the second one involving three fuzzy sets and their pairwise intersections, which state that certain inequalities satisfied by cardinalities of ordinary sets, are preserved under fuzzification.

5.3.1 A meta-theorem for inequalities involving two fuzzy sets

Before we can formulate and prove our meta-theorem, we first have to prove three lemmata.

Lemma 5.1 *Consider $a, b, u, n \in \mathbb{N}$, then the following inequalities are fulfilled*

$$\begin{aligned} 0 \leq a, b \leq n\,, \\ \max(a + b - n, 0) \leq u \leq \min(a, b)\,, \end{aligned} \tag{5.3}$$

if and only if there exist two ordinary subsets A and B of a finite universe X such that $|A| = a$, $|B| = b$, $|A \cap B| = u$ and $|X| = n$.

Proof. The proof from right to left is nothing else but Bell inequality B_2^1. Consider $a, b, u, n \in \mathbb{N}$ satisfying inequalities (5.3), then we need to build two subsets A and B of some universe X. Consider the setting as in Figure 5.1. It suffices to show that there exist $a_1, a_2, c_{12}, d_{12} \in \mathbb{N}$ such that

$$\begin{aligned} &|A| = a = a_1 + c_{12}\,, |A \cap B| = u = c_{12}\,, \\ &|B| = b = a_2 + c_{12}\,, |X| = n = a_1 + a_2 + c_{12} + d_{12}\,. \end{aligned}$$

Due to (5.3), the unique solution $c_{12} := u$, $a_1 := a - c_{12} = a - u$, $a_2 := b - c_{12} = b - u$, $d_{12} := n - a_1 - a_2 + c_{12} = n - a - b + u$ is obviously positive. □

Lemma 5.2 *If for any ordinary subsets A and B of an arbitrary finite universe X it holds that*

$$\mathcal{H}(|A|, |B|, |A \cap B|, |X|) \geq 0\,,$$

where $\mathcal{H}$ denotes a (continuous) function that is homogeneous in its arguments, then it also holds that

$$\mathcal{H}(a, b, u, n) \geq 0\,,$$

for any $a, b, u \in \mathbb{Q}^+$ and $n \in \mathbb{N}$ satisfying inequalities (5.3).

Proof. Consider $a, b, u \in \mathbb{Q}^+$ and $n \in \mathbb{N}$ satisfying inequalities (5.3). Let λ be the lowest common multiple of the denominators of a, b and u, then $a' = \lambda a$, $b' = \lambda b$, $u' = \lambda u$ and $n' = \lambda n$. Moreover, the following inequalities are fulfilled:

$$\begin{aligned} 0 \leq a', b' \leq n'\,, \\ \max(a' + b' - n', 0) \leq u' \leq \min(a', b')\,. \end{aligned}$$

Since $a', b', u', n' \in \mathbb{N}$, according to Lemma 5.1 there exist two subsets A and B of a finite universe X, such that $|A| = a'$, $|B| = b'$, $|A \cap B| = u'$ and $|X| = n'$ and therefore inequality $\mathcal{H}(a', b', u', n') \geq 0$ is fulfilled. Since $\mathcal{H}$ is homogeneous in its arguments, also $\mathcal{H}(a, b, u, n) \geq 0$ holds. □

Lemma 5.3 *If for any ordinary subsets A and B of an arbitrary finite universe X it holds that*

$$\mathcal{H}(|A|, |B|, |A \cap B|, |X|) \geq 0\,,$$

where $\mathcal{H}$ denotes a continuous function that is homogeneous in its arguments, then it also holds that

$$\mathcal{H}(a, b, u, n) \geq 0\,,$$

for any $a, b, u \in \mathbb{R}^+$ and $n \in \mathbb{N}$ satisfying inequalities (5.3).

Proof. Consider $a, b, u \in \mathbb{R}^+$ and $n \in \mathbb{N}$ satisfying (5.3). The idea is to approximate a, b, u within an ϵ-range by $a_\epsilon, b_\epsilon, u_\epsilon \in \mathbb{Q}^+$ such that $a_\epsilon, b_\epsilon, u_\epsilon$ and n also satisfy (5.3). Some boundary cases are easily dealt with.

(i) If $a = 0$ then also $u = 0$, and we consider $b_\epsilon \in [b-\epsilon, b+\epsilon] \cap [0, n]$ for any $\epsilon > 0$ and put $a_\epsilon = u_\epsilon = 0$. The case $b = 0$ is similar.

(ii) If $a = n$ then $b = u$, and we consider $b_\epsilon \in [b-\epsilon, b+\epsilon] \cap [0, n]$ for any $\epsilon > 0$ and put $a_\epsilon = n$ and $u_\epsilon = b_\epsilon$. The case $b = n$ is similar.

(iii) If $a > 0$, $b > 0$ and $u = 0$, then we consider $a_\epsilon \in [a-\epsilon, a]$ and $b_\epsilon \in [b-\epsilon, b]$, for any $\epsilon \leq \min(a, b)$ and put $u_\epsilon = 0$.

From here on, we can assume that $0 < a < n$, $0 < b < n$ and $0 < u$.

(iv) If $a + b - n \leq 0$, then we distinguish two cases.

  (a) If $u = \min(a, b)$, then we consider $a_\epsilon \in [a-\epsilon, a]$ and $b_\epsilon \in [b-\epsilon, b]$, for any $\epsilon \leq \min(a, b)$ and put $u_\epsilon = \min(a_\epsilon, b_\epsilon)$.

  (b) If $u < \min(a, b)$, then we consider $a_\epsilon \in [a-\epsilon, a]$, $b_\epsilon \in [b-\epsilon, b]$ and $u_\epsilon \in [u-\epsilon, u]$, for any $\epsilon \leq \min(a, b, u, \min(a, b) - u)$.

(v) Consider $a + b - n > 0$. Since $a + b - n = u = \min(a, b)$ implies $\max(a, b) = n$, only three cases remain to be studied.

  (a) If $a + b - n = u < \min(a, b)$, we consider $a_\epsilon \in [a - \frac{\epsilon}{2}, a]$ and $b_\epsilon \in [b - \frac{\epsilon}{2}, b]$, for $\epsilon \leq 2\min(a, b)$ and put $u_\epsilon = \max(a_\epsilon + b_\epsilon - n, 0)$.

  (b) If $a + b - n < u = \min(a, b)$, we consider $a_\epsilon \in [a-\epsilon, a]$, $b_\epsilon \in [b-\epsilon, b]$, for $\epsilon \leq \min(a, b)$ and put $u_\epsilon = \min(a_\epsilon, b_\epsilon)$.

  (c) If $a + b - n < u < \min(a, b)$, we consider $a_\epsilon \in [a-\epsilon, a]$, $b_\epsilon \in [b-\epsilon, b]$ and $u_\epsilon \in [u-\epsilon, u]$, for $\epsilon \leq \min(a, b, u, \min(a, b) - u, u - a - b + n)$.

One easily verifies that in all cases a_ϵ, b_ϵ and u_ϵ are positive and $|a - a_\epsilon| \leq \epsilon$, $|b - b_\epsilon| \leq \epsilon$ and $|u - u_\epsilon| \leq \epsilon$. Using Lemma 5.2 it follows that $\mathcal{H}(a_\epsilon, b_\epsilon, u_\epsilon, n) \geq 0$. Since $\mathcal{H}$ is continuous, considering the limit for $\epsilon \to 0$, we can conclude that also the inequality $\mathcal{H}(a, b, u, n) \geq 0$ is fulfilled. □

We are now able to formulate and prove our meta-theorem.

Theorem 5.1 *If for any ordinary subsets A and B of an arbitrary finite universe X it holds that*

$$\mathcal{H}(|A|, |B|, |A \cap B|, |X|) \geq 0\,, \tag{5.4}$$

where $\mathcal{H}$ denotes a continuous function that is homogeneous in its arguments, then it also holds for any fuzzy sets on an arbitrary finite universe Y, provided the commutative conjunctor I modelling fuzzy set intersection satisfies Bell inequality I_2^1.

Proof. Consider two fuzzy sets C and D on an arbitrary finite universe Y. Let $n := |Y|$ and $Y = \{y_1, \ldots, y_n\}$, $a := |C|$, $b := |D|$ and $u := |C \cap D|$. Then obviously $0 \leq a, b \leq n$. Since any conjunctor I is bounded from above by $T_\mathbf{M}$ it holds that

$$u = \sum_{i=1}^{n} I(C(y_i), D(y_i)) \leq \sum_{i=1}^{n} \min(C(y_i), D(y_i)) \leq \min(a, b)\,.$$

If I is also bounded from below by $T_\mathbf{L}$, i.e. if I satisfies inequality I_2^1, we can also conclude that

$$u \geq \sum_{i=1}^{n} \max(C(y_i) + D(y_i) - 1, 0) \geq a + b - n\,.$$

Since $a, b, u \in \mathbb{R}^+$ and $n \in \mathbb{N}$ satisfy inequalities (5.3), we can use Lemma 5.3 to conclude that inequality (5.4) is also fulfilled for C and D on Y. □

Having a closer look at the proof of Theorem 5.1, we can derive the following, more specialised version of the meta-theorem, which does not require Bell inequality I_2^1.

Theorem 5.2 *Under the assumptions of Theorem 5.1: if $\mathcal{H}$ does not depend explicitly upon $|X|$, then (5.4) also holds for any fuzzy sets on an arbitrary finite universe Y.*

Proof. Since $\mathcal{H}$ does not depend upon $|X|$, we can augment the value of n (i.e. we add zero components to the fuzzy sets involved) such that the inequality $a + b - n \leq u$ is always satisfied without changing the value of a, b and u. □

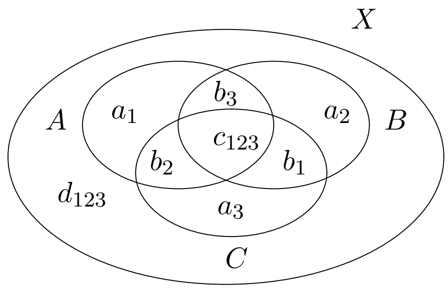


Figure 5.2: Notations for cardinalities associated with three ordinary subsets A, B and C of a finite universe X.

5.3.2 A meta-theorem for inequalities involving three fuzzy sets

The same reasoning as in the previous subsection can be followed for inequalities involving three sets and their pairwise intersections.

Lemma 5.4 *Consider a, b, c, u, v, w, $n \in \mathbb{N}$, then the following inequalities are fulfilled*

$$\begin{aligned} 0 \leq a, b, c \leq n\,, \\ \max(a+b-n,0) \leq u \leq \min(a,b)\,, \\ \max(a+c-n,0) \leq v \leq \min(a,c)\,, \\ \max(b+c-n,0) \leq w \leq \min(b,c)\,, \\ w-u-v+a \geq 0\,, \\ v-u-w+b \geq 0\,, \\ u-v-w+c \geq 0\,, \\ n-(a+b+c-u-v-w) \geq 0\,, \end{aligned} \tag{5.5}$$

if and only if there exist three ordinary subsets A, B and C of a finite universe X such that $|A| = a$, $|B| = b$, $|C| = c$, $|X| = n$, $|A \cap B| = u$, $|A \cap C| = v$ and $|B \cap C| = w$.

Proof. The proof from right to left is nothing else but Bell inequalities B_2^1–B_3^3. Consider a, b, c, u, v, w, $n \in \mathbb{N}$ satisfying inequalities (5.5), then we need to build three subsets A, B and C of some universe X. Consider the setting as in Figure 5.2. It suffices to show that there exist

$a_1, a_2, a_3, b_1, b_2, b_3, c_{123}, d_{123} \in \mathbb{N}$ such that

$$\begin{array}{ll} |A| = a = a_1 + b_2 + b_3 + c_{123}\,, & |A \cap B| = u = b_3 + c_{123}\,, \\ |B| = b = a_2 + b_1 + b_3 + c_{123}\,, & |A \cap C| = v = b_2 + c_{123}\,, \\ |C| = c = a_3 + b_1 + b_2 + c_{123}\,, & |B \cap C| = w = b_1 + c_{123} \end{array}$$

and

$$|X| = n = a_1 + a_2 + a_3 + b_1 + b_2 + b_3 + c_{123} + d_{123}\,.$$

We obtain the following solution

$$\begin{array}{ll} a_1 := a - u - v + c_{123}\,, & b_1 := w - c_{123}\,, \\ a_2 := b - u - w + c_{123}\,, & b_2 := v - c_{123}\,, \\ a_3 := c - v - w + c_{123}\,, & b_3 := u - c_{123}\,, \end{array}$$

and

$$d_{123} := n - (a + b + c - u - v - w) - c_{123}\,,$$

with c_{123} still to be chosen. Due to inequalities (5.5), it holds that

$$\begin{array}{l} p := \max(u + v - a, u + w - b, v + w - c, 0) \\ \leq q := \min(u, v, w, n - (a + b + c - u - v - w))\,. \end{array}$$

Choosing any $c_{123} \in [p, q]$ renders the solution positive and concludes the proof. □

Lemma 5.5 *If for any ordinary subsets A, B and C of an arbitrary finite universe X it holds that*

$$\mathcal{H}(|A|, |B|, |C|, |A \cap B|, |A \cap C|, |B \cap C|, |X|) \geq 0\,,$$

where $\mathcal{H}$ denotes a (continuous) function that is homogeneous in its arguments, then it also holds that

$$\mathcal{H}(a, b, c, u, v, w, n) \geq 0\,,$$

for any $a, b, c, u, v, w \in \mathbb{Q}^+$ and $n \in \mathbb{N}$ satisfying inequalities (5.5).

Proof. Similar to the proof of Lemma 5.2. □

Lemma 5.6 *If for any ordinary subsets A, B and C of an arbitrary finite universe X it holds that*

$$\mathcal{H}(|A|, |B|, |C|, |A \cap B|, |A \cap C|, |B \cap C|, |X|) \geq 0\,,$$

where $\mathcal{H}$ denotes a continuous function that is homogeneous in its arguments, then it also holds that

$$\mathcal{H}(a, b, c, u, v, w, n) \geq 0\,,$$

for any $a, b, c, u, v, w \in \mathbb{R}^+$ and $n \in \mathbb{N}$ satisfying inequalities (5.5).

Proof. Similar to the proof of Lemma 5.3. □

We are now ready to formulate and prove our second meta-theorem.

Theorem 5.3 *If for any ordinary subsets A, B and C of an arbitrary finite universe X it holds that*

$$\mathcal{H}(|A|, |B|, |C|, |A \cap B|, |A \cap C|, |B \cap C|, |X|) \geq 0\,, \tag{5.6}$$

where $\mathcal{H}$ denotes a continuous function that is homogeneous in its arguments, then it also holds for any fuzzy sets on an arbitrary finite universe Y, provided the commutative conjunctor I modelling fuzzy set intersection satisfies Bell inequalities I_3^2 and I_3^3.

Proof. Consider three fuzzy sets D, E and F on an arbitrary finite universe Y. Let $n := |Y|$, $a := |D|$, $b := |E|$, $c := |F|$, $u := |D \cap E|$, $v := |D \cap F|$ and $w := |E \cap F|$. Then obviously $0 \leq a, b, c \leq n$.

Since any conjunctor I is bounded from above by $T_{\mathbf{M}}$ it again holds that $u \leq \min(a, b)$, $v \leq \min(a, c)$ and $w \leq \min(b, c)$. If I is also bounded from below by $T_{\mathbf{L}}$, i.e. if I satisfies inequality I_2^1 (which follows from I_3^2), it also holds that $a+b-n \leq u$, $a+c-n \leq v$ and $b+c-n \leq w$. Moreover, if I satisfies Bell inequality I_3^2, we can conclude that $w - u - v + a \geq 0$, $v - u - w + b \geq 0$ and $u - v - w + c \geq 0$. In the same way, if I satisfies Bell inequality I_3^3, the inequality $n - (a + b + c - u - v - w) \geq 0$ follows.

Since $a, b, u \in \mathbb{R}^+$ and $n \in \mathbb{N}$ satisfy inequalities (5.5), we can use Lemma 5.6 to conclude that inequality (5.6) is also fulfilled for D, E and F on Y. □

Again, a more specialised version of this meta-theorem can be stated.

Theorem 5.4 *Under the assumptions of Theorem 5.3: if $\mathcal{H}$ does not depend explicitly upon $|X|$, then (5.6) also holds for any fuzzy sets on an arbitrary finite universe Y, provided the commutative conjunctor I satisfies Bell inequality I_3^2.*

Proof. Since $\mathcal{H}$ does not depend upon $|X|$, we can augment the value of n such that the inequalities $a + b - n \leq u$, $a + c - n \leq v$, $b + c - n \leq w$ and $n - (a + b + c - u - v - w) \geq 0$ are always satisfied without changing the value of a, b, c, u, v and w. □

5.4 Applications

5.4.1 Linear inequalities involving two sets

Since the Bell inequalities B_2^1–B_3^3 are particular linear inequalities, we wonder which linear inequalities are valid in general, both in the crisp case and in the fuzzy case.

Theorem 5.5 *Consider a linear inequality with real coefficients of the form*

$$a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \geq 0\,, \tag{5.7}$$

with A and B ordinary subsets of an arbitrary finite universe X.

The following statements are equivalent:

(i) Inequality (5.7) holds for all ordinary subsets A and B of X.

(ii) Inequality (5.7) holds for all $(A, B) \in \{\emptyset, X\}^2$.

(iii) The following conditions on a_1, a_2, b_{12} and c hold:

$$\begin{aligned} c &\geq 0\,, \\ a_1 + c &\geq 0\,, \\ a_2 + c &\geq 0\,, \\ a_1 + a_2 + b_{12} + c &\geq 0\,. \end{aligned} \tag{5.8}$$

Proof. We give a circular proof. Since (i) trivially implies (ii) and it is easily verified that (iii) follows from (ii), it suffices to show that (iii) implies (i). Consider two ordinary subsets A and B of $X = \{x_1, \ldots, x_n\}$. Identifying A and B with their characteristic mapping, inequality (5.7) is clearly equivalent with

$$\sum_{i=1}^{n} (a_1 A(x_i) + a_2 B(x_i) + b_{12} \min(A(x_i), B(x_i)) + c) \geq 0\,.$$

Due to conditions (5.8), every term in the above sum is positive, and therefore inequality (5.7) is fulfilled as well. □

We next apply our first meta-theorem to these linear inequalities.

Theorem 5.6 *Consider a linear inequality with real coefficients of the form*

$$a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \geq 0\,, \tag{5.9}$$

with A and B fuzzy sets on an arbitrary finite universe X.

(i) If I satisfies I_2^1, then inequality (5.9) holds for all fuzzy sets A and B on X if and only if conditions (5.8) hold.

(ii) If $c = 0$, then inequality (5.9) holds for all fuzzy sets A and B on X if and only if conditions (5.8) hold.

Proof. The first part follows from Theorems 5.1 and 5.5, while the second part follows from Theorems 5.2 and 5.5. □

The following theorem shows that in some cases more general results can be obtained through a direct proof. Indeed, in some cases I does not need to satisfy I_2^1, i.e. it does not need to be bounded from below by $T_{\mathbf{L}}$.

Theorem 5.7 *Consider a linear inequality with real coefficients of the form (5.9), with A and B fuzzy sets on an arbitrary finite universe X, such that* $\max(a_1, a_2) \geq 0$ *or* $b_{12} \leq 0$. *Then inequality (5.9) holds for all fuzzy sets A and B on X if and only if conditions (5.8) hold.*

Proof. This theorem, in particular the fact that I does not need to satisfy I_2^1, does not follow from the meta-theorem (except for the case $c = 0$) and needs to be proven explicitly.

Suppose that conditions (5.8) hold. A direct proof of Theorem 5.5 (which is given in Section 5.4.3) shows that only in the case $a_1 < 0$, $a_2 < 0$ and $b_{12} > 0$ (and then necessarily $c > 0$) the left-hand side of B_2^1 (or equivalently, the left-hand side of I_2^1) needs to be invoked. Indeed, consider for instance the subcase $\min(|a_1|, |a_2|, b_{12}) = b_{12}$ (the other subcases being similar), then we can write $a_1 = a'_1 - b_{12}$ and $a_2 = a'_2 - b_{12}$, with $a'_1 \leq 0$ and $a'_2 \leq 0$. For any two fuzzy sets A and B on $X = \{x_1, \ldots, x_n\}$, it then holds that

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \\
&= a'_1|A| + a'_2|B| + b_{12}(-|A| - |B| + |A \cap B|) + c\,n\,.
\end{aligned}$$

Using the left-hand side of B_2^1 it follows that

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \\
&\geq a'_1|A| + a'_2|B| - b_{12}\,n + c\,n \\
&\geq a'_1\,n + a'_2\,n - b_{12}\,n + c\,n \\
&= a_1\,n + a_2\,n + b_{12}\,n + c\,n \geq 0\,.
\end{aligned}$$

The positivity in the last step is due to the last of conditions (5.8).

Let us demonstrate one of the other cases where the left-hand side of B_2^1 is not necessary. Suppose that $a_1 \geq 0$, $a_2 \geq 0$ and $b_{12} < 0$. Consider for instance the subcase $\min(a_1, a_2, |b_{12}|) = |b_{12}|$, then we can write $a_1 = a_1' - b_{12}$ and $a_2 = a_2' - b_{12}$, with $a_1' \geq 0$ and $a_2' \geq 0$. For any two fuzzy sets A and B on $X = \{x_1, \ldots, x_n\}$, it then holds that

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c\,n \\
&= -b_{12}(|A| + |B| - |A \cap B|) + a_1'|A| + a_2'|B| + c\,n\,.
\end{aligned}$$

Using the right-hand side of B_2^1 it follows that $|A| + |B| - |A \cap B| \geq 0$ and hence

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c\,n \\
&\geq a_1'|A| + a_2'|B| + c\,n \geq 0\,.
\end{aligned}$$

This completes the proof. $\square$

5.4.2 Linear inequalities involving three sets

Theorem 5.8 *Consider a linear inequality with real coefficients of the form*

$$a_1|A|+a_2|B|+a_3|C|+b_{12}|A\cap B|+b_{13}|A\cap C|+b_{23}|B\cap C|+c|X| \geq 0 \quad (5.10)$$

with A, B and C ordinary subsets of an arbitrary finite universe X.

The following statements are equivalent:

(i) Inequality (5.10) holds for all ordinary subsets A, B and C of X.

(ii) Inequality (5.10) holds for all $(A, B, C) \in \{\emptyset, X\}^3$ of X.

(iii) The following conditions on a_i, b_{ij} and c hold for all $i = 1, \ldots, 3$ and $j > i$:

$$\begin{aligned}
c &\geq 0\,, \\
a_i + c &\geq 0\,, \\
a_i + a_j + b_{ij} + c &\geq 0\,, \\
a_1 + a_2 + a_3 + b_{12} + b_{13} + b_{23} + c &\geq 0\,. \qquad (5.11)
\end{aligned}$$

Proof. We give a circular proof. Since (i) trivially implies (ii) and it is easily verified that (iii) follows from (ii), it suffices to show that (iii) implies (i). Consider three ordinary subsets A, B and C of $X =$

$\{x_1, \ldots, x_n\}$. Identifying A, B and C with their characteristic mapping, inequality (5.10) is clearly equivalent with

$$\sum_{i=1}^{n} (a_1 A(x_i) + a_2 B(x_i) + a_3 C(x_i) + b_{12} \min(A(x_i), B(x_i)) \\ + b_{13} \min(A(x_i), C(x_i)) + b_{23} \min(B(x_i), C(x_i)) + c) \geq 0 \,. \quad (5.12)$$

Due to conditions (5.11), every term in the above sum is positive, and therefore inequality (5.10) is fulfilled as well. □

Now, we can apply our second meta-theorem.

Theorem 5.9 *Consider a linear inequality with real coefficients of the form*

$$a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| + b_{23}|B \cap C| + c|X| \geq 0 \,, \quad (5.13)$$

with A, B and C fuzzy sets on an arbitrary finite universe X.

(i) If I satisfies I_3^2 and I_3^3, then inequality (5.13) holds for all fuzzy sets A, B and C on X if and only if conditions (5.11) hold.

(ii) If $c = 0$ and I satisfies I_3^2, then inequality (5.13) holds for all fuzzy sets A, B and C on X if and only if conditions (5.11) hold.

Proof. The first part follows from Theorems 5.3 and 5.8, while the second part follows from Theorems 5.4 and 5.8. □

The following theorem shows that in some cases more general results can be obtained through a direct proof. Indeed, in some cases I does not need to satisfy I_3^3.

Theorem 5.10 *Consider a linear inequality with real coefficients of the form (5.13), with A, B and C fuzzy sets on an arbitrary finite universe X, such that*

$$\begin{aligned} & \max(a_i, a_j, a_k) \geq 0 \ \vee \ \min(b_{ij}, b_{jk}, b_{ik}) \leq 0 \\ \vee \ & [\ (a_j + b_{ij} > 0 \ \vee \ c + a_k - b_{ij} \geq 0) \\ & \wedge (a_j + b_{ij} \leq 0 \ \vee \ a_i \leq a_j + a_k \ \vee \ a_i + b_{jk} \leq a_j \ \vee \ c + a_i + a_k \geq 0)\ ] \,, \end{aligned}$$

and I satisfies I_3^2. Then inequality (5.13) holds for all fuzzy sets A, B and C on X if and only if conditions (5.11) hold.

Proof. This theorem, in particular the fact that I does not need to satisfy I_3^3, does not follow from the meta-theorem (except for the case $c = 0$) and needs to be proven explicitly.

Suppose that conditions (5.11) hold. A direct proof of Theorem 5.8 (which is given in Section 5.4.3) shows that only in the case

$$\begin{aligned}
&\max(a_i, a_j, a_k) < 0 \ \wedge \ \min(b_{ij}, b_{jk}, b_{ik}) > 0 \\
\wedge \ &[\ (a_j + b_{ij} \leq 0 \ \wedge \ c + a_k - b_{ij} < 0) \\
&\vee (a_j + b_{ij} > 0 \ \wedge \ a_i > a_j + a_k \ \wedge \ a_i + b_{jk} > a_j \ \wedge \ c + a_i + a_k < 0)\ ]\ ,
\end{aligned}$$

inequality B_3^3 (or equivalently, inequality I_3^3) needs to be invoked.

Indeed, consider for instance the subcase $a_i < 0$, $b_{ij} \geq 0$, $\min(|a_1|, |a_2|, |a_3|, b_{12}, b_{13}, b_{23}) = b_{12}$ and $c - b_{12} + a_3 < 0$. Then, we can write $a_1 = a_1' - b_{12}$, $a_2 = a_2' - b_{12}$, $a_3 = a_3' - b_{12}$, $b_{13} = b_{13}' + b_{12}$ and $b_{23} = b_{23}' + b_{12}$, with $a_1' \leq 0$, $a_2' \leq 0$, $a_3' \leq 0$, $b_{13}' \geq 0$ and $b_{23}' \geq 0$. We then obtain

$$\begin{aligned}
&a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| + b_{23}|B \cap C| + c|X| \\
&= a_1'|A| + a_2'|B| + a_3'|C| + b_{13}'|A \cap C| + b_{23}'|B \cap C| + c\,n \\
&\quad + \ b_{12}(-|A| - |B| - |C| + |A \cap B| + |A \cap C| + |B \cap C|) \\
&\geq a_1'|A| + a_2'|B| + a_3'|C| + b_{13}'|A \cap C| + b_{23}'|B \cap C| + (c - b_{12})\,n\,.
\end{aligned}$$

Furthermore, suppose $\min(|a_1'|, |a_3'|, b_{13}) = |a_3'|$ (again, other cases can be proven in a similar way), then we can write: $a_1' = a_1'' + a_3'$ and $b_{13}' = b_{13}'' - a_3'$, with $a_1'' \leq 0$ and $b_{13}'' \geq 0$. We obtain:

$$\begin{aligned}
&a_1'|A| + a_2'|B| + a_3'|C| + b_{13}'|A \cap C| + b_{23}'|B \cap C| + (c - b_{12})\,n \\
&\geq a_1''|A| + a_2'|B| + b_{13}''|A \cap C| + b_{23}'|B \cap C| + (c - b_{12} + a_3')\,n \\
&\geq a_1''\,n + a_2'\,n + b_{13}''|A \cap C| + b_{23}'|B \cap C| + (c - b_{12} + a_3')\,n \\
&\geq 0\,,
\end{aligned}$$

since $c - b_{12} + a_3' + a_1'' + a_2' = c + a_1 + a_2 + b_{12} \geq 0$.

Let us demonstrate one of the other cases where inequality B_3^3 is not necessary. Suppose that $a_i < 0$, $b_{ij} \geq 0$, $\min(|a_1|, |a_2|, |a_3|, b_{12}, b_{13}, b_{23}) = b_{12}$ and $c - b_{12} + a_3 \geq 0$. Then, we can write $a_1 = a_1' - b_{12}$ and $a_2 = a_2' - b_{12}$, with $a_1' \leq 0$ and $a_2' \leq 0$. We then obtain

$$\begin{aligned}
&a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| + b_{23}|B \cap C| + c\,n \\
&= a_1'|A| + a_2'|B| + a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| + c\,n \\
&\quad + \ b_{12}(-|A| - |B| + |A \cap B|) \\
&\geq a_1'|A| + a_2'|B| + a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| + (c - b_{12})\,n\,.
\end{aligned}$$

Furthermore, suppose $\min(|a'_2|, |a_3|, b_{23}) = |a'_2|$ (again, other subcases can be proven in a similar way), then we can write: $a_3 = a'_3 + a'_2$ and $b_{23} = b'_{23} - a'_2$, with $a'_3 \leq 0$ and $b'_{23} \geq 0$. We obtain:

$$\begin{aligned} & a'_1|A| + a'_2|B| + a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| + (c - b_{12})\, n \\ & \geq a'_1|A| + a'_3|C| + b_{13}|A \cap C| + b'_{23}|B \cap C| + (c - b_{12} + a'_2)\, n\,. \end{aligned}$$

Consider the subcase $\min(|a'_1|, |a'_3|, b_{13}) = |a'_1|$, then we can write: $a'_3 = a''_3 + a'_1$ and $b_{13} = b'_{13} - a'_1$, with $a''_3 \leq 0$ and $b'_{13} \geq 0$. We obtain:

$$\begin{aligned} & a'_1|A| + a'_3|C| + b_{13}|A \cap C| + b'_{23}|B \cap C| + (c - b_{12} + a'_2)\, n \\ & \geq a''_3|C| + b'_{13}|A \cap C| + b'_{23}|B \cap C| + (c - b_{12} + a'_1 + a'_2)\, n \\ & \geq (c - b_{12} + a'_1 + a'_2 + a''_3)\, n \\ & \geq 0\,, \end{aligned}$$

since $c - b_{12} + a'_1 + a'_2 + a''_3 = c - b_{12} + a_3 \geq 0$.
This completes the proof. □

5.4.3 Direct proofs of Theorems 5.5 and 5.8

For the sake of completeness, we give the direct proofs of Theorem 5.5 and Theorem 5.8.

Theorem 5.11 *A linear inequality with real coefficients of the form*

$$a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \geq 0 \tag{5.14}$$

holds for all ordinary subsets A and B of a finite universe X if and only if it holds for all $(A, B) \in \{\emptyset, X\}^2$.

Proof. Suppose that inequality (5.14) holds for any $(A, B) \in \{\emptyset, X\}^2$, then the following conditions on a_1, a_2, b_{12} and c hold:

$$\begin{aligned} c &\geq 0\,, \\ a_1 + c &\geq 0\,, \\ a_2 + c &\geq 0\,, \\ a_1 + a_2 + b_{12} + c &\geq 0\,. \end{aligned}$$

Consider two arbitrary subsets A and B of X. We split the proof into several cases. Suppose $a_1 \geq 0$ and $a_2 \geq 0$. If $b_{12} \geq 0$, then inequality (5.14) is trivially fulfilled. Suppose $b_{12} < 0$ and consider the following subcases:

(i) The case $\min(a_1, a_2, |b_{12}|) = |b_{12}|$. Then we rewrite $a_1 = a'_1 - b_{12}$ and $a_2 = a'_2 - b_{12}$, with $a'_1 \geq 0$ and $a'_2 \geq 0$. Due to the right-hand side of B_2^1, it holds that

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \\
&= -b_{12}(|A| + |B| - |A \cap B|) + a'_1|A| + a'_2|B| + c|X| \\
&\geq a'_1|A| + a'_2|B| + c|X| \geq 0\,.
\end{aligned}$$

(ii) The case $\min(a_1, a_2, |b_{12}|) = a_1$ (the case $\min(a_1, a_2, |b_{12}|) = a_2$ is analogous). Then we rewrite $b_{12} = b'_{12} - a_1$, with $b'_{12} \leq 0$. Due to the right-hand side of B_2^1, it holds that

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \\
&= a_1(|A| - |A \cap B|) + a_2|B| + b'_{12}|A \cap B| + c|X| \\
&\geq a_2|B| + b'_{12}|A \cap B| + c|X|\,.
\end{aligned}$$

Consider the subcase $\min(a_2, |b'_{12}|) = |b'_{12}|$ (the subcase $\min(a_2, |b_{12}|) = a_2$ is analogous). We rewrite $b'_{12} = b''_{12} - a_2$, with $b''_{12} \leq 0$, and obtain

$$\begin{aligned}
&a_2|B| + b'_{12}|A \cap B| + c|X| \\
&= a_2(|B| - |A \cap B|) + b''_{12}|A \cap B| + c|X| \\
&\geq b''_{12}|A \cap B| + c|X| \\
&\geq (b''_{12} + c)|X| \geq 0\,,
\end{aligned}$$

since $b''_{12} + c = a_1 + a_2 + b_{12} + c \geq 0$.

In the same way, one can prove that inequality (5.14) holds for the cases $a_1 \geq 0$, $a_2 \leq 0$; $a_1 \leq 0$, $a_2 \geq 0$ and $a_1 \leq 0$, $a_2 \leq 0$. Nevertheless, we mention here the proof for the case $a_1 < 0$, $a_2 < 0$ and $b_{12} > 0$, since it is the only case where we need to invoke the left-hand side of B_1. Suppose $\min(|a_1|, |a_2|, b_{12}) = b_{12}$ (the other cases being similar), then we write $a_1 = a'_1 - b_{12}$ and $a_2 = a'_2 - b_{12}$, with $a'_1 \leq 0$ and $a'_2 \leq 0$. We obtain the following

$$\begin{aligned}
&a_1|A| + a_2|B| + b_{12}|A \cap B| + c|X| \\
&= a'_1|A| + a'_2|B| + b_{12}(-|A| - |B| + |A \cap B|) + c|X| \\
&\geq a'_1|A| + a'_2|B| - b_{12}|X| + c|X| \\
&\geq (a'_1 + a'_2 - b_{12} + c)|X| \geq 0\,,
\end{aligned}$$

since $a'_1 + a'_2 - b_{12} + c = a_1 + a_2 + b_{12} + c \geq 0$.
This completes our proof. $\square$

Theorem 5.12 *A linear inequality with real coefficients of the form:*

$$a_1|A|+a_2|B|+a_3|C|+b_{12}|A\cap B|+b_{13}|A\cap C|+b_{23}|B\cap C|+c|X| \geq 0 \quad (5.15)$$

holds for all ordinary subsets A, B and C of a finite universe X if and only if it holds for all $(A,B,C) \in \{\emptyset, X\}^3$.

Proof. Suppose that inequality (5.15) holds for any $(A,B,C) \in \{\emptyset, X\}^3$, then the following conditions on a_i, b_{ij} and c hold for all $i = 1,\ldots,3$ and $j > i$:

$$\begin{aligned} c &\geq 0\,,\\ a_i + c &\geq 0\,,\\ a_i + a_j + b_{ij} + c &\geq 0\,,\\ a_1 + a_2 + a_3 + b_{12} + b_{13} + b_{23} + c &\geq 0\,. \end{aligned}$$

Again, we split up the proof into several parts. Suppose all $a_i \geq 0$. If also all $b_{ij} \geq 0$, inequality (5.15) is trivially fulfilled.

(i) Suppose $b_{12} \leq 0$, $b_{13} \geq 0$ and $b_{23} \geq 0$. Since $a_1|A| + a_2|B| + b_{12}|A \cap B| + c' \geq 0$ for any a_1, a_2, b_{12} and c satisfying $c \geq 0$, $a_1 + c \geq 0$, $a_2 + c \geq 0$, and $a_1 + a_2 + b_{12} + c \geq 0$ (see Theorem 5.11), we can conclude that

$$\begin{aligned} & a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C|\\ & +\ b_{23}|B \cap C| + c|X|\\ & \geq a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| \geq 0\,. \end{aligned}$$

(ii) Next, suppose $b_{12} \leq 0$, $b_{13} \leq 0$ and $b_{23} \geq 0$. Furthermore, suppose $b_{23} \leq \min(a_1, |b_{12}|, |b_{13}|)$, then we can rewrite $a_1 = a_1' + b_{23}$, $b_{12} = b_{12}' - b_{23}$ and $b_{13} = b_{13}' - b_{23}$, with $a_1' \geq 0$, $b_{12}' \leq 0$ and $b_{13}' \leq 0$. Due to B_2, it holds that

$$\begin{aligned} & a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C|\\ & +\ b_{23}|B \cap C| + c|X|\\ & = b_{23}(|A| - |A \cap B| - |A \cap C| + |B \cap C|)\\ & \quad +\ a_1'|A| + a_2|B| + a_3|C| + b_{12}'|A \cap B| + b_{13}'|A \cap C| + c|X|\\ & \geq a_1'|A| + a_2|B| + a_3|C| + b_{12}'|A \cap B| + b_{13}'|A \cap C| + c|X|\,. \end{aligned}$$

Furthermore, consider the subcase $\min(a_2, |b_{12}'|) = |b_{12}'|$ (other subcases can be proven in a similar way). Then we write $a_2 =$

$a'_2 - b'_{12}$, with $a'_2 \geq 0$. Due to the right-hand side of B_1, it follows that

$$\begin{aligned} & a'_1|A| + a_2|B| + a_3|C| + b'_{12}|A \cap B| + b'_{13}|A \cap C| + c|X| \\ & \geq a'_1|A| + a'_2|B| + a_3|C| + b'_{13}|A \cap C| + c|X| \geq 0\,, \end{aligned}$$

since $a'_1|A| + a_3|C| + b'_{13}|A \cap C| + c' \geq 0$, due to Theorem 5.11.

All other cases and subcases are proven in a similar way. Nevertheless, we mention here the proof for some subcases where we need to invoke B_3. Suppose all $a_i < 0$ and all $b_{ij} \geq 0$.

(i) Suppose $\min(|a_1|, |a_2|, |a_3|, b_{12}, b_{13}, b_{23}) = b_{12}$ and $c - b_{12} + a_3 < 0$. Then, we can write $a_1 = a'_1 - b_{12}$, $a_2 = a'_2 - b_{12}$, $a_3 = a'_3 - b_{12}$, $b_{13} = b'_{13} + b_{12}$ and $b_{23} = b'_{23} + b_{12}$, with $a'_1 \leq 0$, $a'_2 \leq 0$, $a'_3 \leq 0$, $b'_{13} \geq 0$ and $b'_{23} \geq 0$. We then obtain

$$\begin{aligned} & a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| \\ & + b_{23}|B \cap C| + c|X| \\ & = a'_1|A| + a'_2|B| + a'_3|C| + b'_{13}|A \cap C| + b'_{23}|B \cap C| + c|X| \\ & \quad + \ b_{12}(-|A| - |B| - |C| + |A \cap B| + |A \cap C| + |B \cap C|) \\ & \geq a'_1|A| + a'_2|B| + a'_3|C| + b'_{13}|A \cap C| + b'_{23}|B \cap C| \\ & \quad + \ (c - b_{12})|X|\,. \end{aligned}$$

Furthermore, suppose $\min(|a'_1|, |a'_3|, b_{13}) = |a'_3|$ (again, other subcases can be proven in a similar way), then we can write: $a'_1 = a''_1 + a'_3$ and $b'_{13} = b''_{13} - a'_3$, with $a''_1 \leq 0$ and $b''_{13} \geq 0$. We obtain:

$$\begin{aligned} & a'_1|A| + a'_2|B| + a'_3|C| + b'_{13}|A \cap C| + b'_{23}|B \cap C| + (c - b_{12})|X| \\ & \geq a''_1|A| + a'_2|B| + b''_{13}|A \cap C| + b'_{23}|B \cap C| + (c - b_{12} + a'_3)|X| \\ & \geq a''_1\, n + a'_2\, n + b''_{13}|A \cap C| + b'_{23}|B \cap C| + (c - b_{12} + a'_3)|X| \\ & \geq 0\,, \end{aligned}$$

since $c - b_{12} + a'_3 + a''_1 + a'_2 = c + a_1 + a_2 + b_{12} \geq 0$.

(i′) Let us demonstrate that when $\min(|a_1|, |a_2|, |a_3|, b_{12}, b_{13}, b_{23}) = b_{12}$ and $c - b_{12} + a_3 \geq 0$ (other subcases can be proven in a similar way), inequality B_3 is not necessary. In that case, we can write $a_1 = a'_1 - b_{12}$ and $a_2 = a'_2 - b_{12}$, with $a'_1 \leq 0$ and $a'_2 \leq 0$. We then obtain

$$\begin{aligned}
&a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| \\
&+ b_{23}|B \cap C| + c|X| \\
&= a_1'|A| + a_2'|B| + a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| + c|X| \\
&\quad + \ b_{12}(-|A| - |B| + |A \cap B|) \\
&\geq a_1'|A| + a_2'|B| + a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| \\
&\quad + \ (c - b_{12})|X| \, .
\end{aligned}$$

Furthermore, suppose $\min(|a_2'|, |a_3|, b_{23}) = |a_2'|$ (again, other subcases can be proven in a similar way), then we can write: $a_3 = a_3' + a_2'$ and $b_{23} = b_{23}' - a_2'$, with $a_3' \leq 0$ and $b_{23}' \geq 0$. We obtain:

$$\begin{aligned}
&a_1'|A| + a_2'|B| + a_3|C| + b_{13}|A \cap C| + b_{23}|B \cap C| + (c - b_{12})|X| \\
&\geq a_1'|A| + a_3'|C| + b_{13}|A \cap C| + b_{23}'|B \cap C| + (c - b_{12} + a_2')|X| \, .
\end{aligned}$$

Consider the subcase $\min(|a_1'|, |a_3'|, b_{13}) = |a_1'|$, then we can write: $a_3' = a_3'' + a_1'$ and $b_{13} = b_{13}' - a_1'$, with $a_3'' \leq 0$ and $b_{13}' \geq 0$. We obtain:

$$\begin{aligned}
&a_1'|A| + a_3'|C| + b_{13}|A \cap C| + b_{23}'|B \cap C| + (c - b_{12} + a_2')|X| \\
&\geq a_3''|C| + b_{13}'|A \cap C| + b_{23}'|B \cap C| + (c - b_{12} + a_1' + a_2')|X| \\
&\geq (c - b_{12} + a_1' + a_2' + a_3'')|X| \\
&\geq 0 \, ,
\end{aligned}$$

since $c - b_{12} + a_1' + a_2' + a_3'' = c - b_{12} + a_3 \geq 0$.

(ii) Suppose $\min(|a_1|, |a_2|, |a_3|, b_{12}, b_{13}, b_{23}) = |a_1|$ and $c + a_1 + a_3 < 0$ (other subcases can be proven in a similar way). Then, we can write $a_2 = a_2' + a_1$, $a_3 = a_3' + a_1$, $b_{12} = b_{12}' - a_1$, $b_{13} = b_{13}' - a_1$ and $b_{23} = b_{23}' - a_1$, with $a_2' \leq 0$, $a_3' \leq 0$, $b_{12}' \geq 0$, $b_{13}' \geq 0$ and $b_{23}' \geq 0$. We then obtain

$$\begin{aligned}
&a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| \\
&+ \ b_{23}|B \cap C| + c|X| \\
&= a_2'|B| + a_3'|C| + +b_{12}'|A \cap B| + b_{13}'|A \cap C| + b_{23}'|B \cap C| + c|X| \\
&\quad + \ a_1(|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|) \\
&\geq a_2'|B| + a_3'|C| + b_{12}'|A \cap B| + b_{13}'|A \cap C| + b_{23}'|B \cap C| \\
&\quad + \ (c + a_1)|X| \, .
\end{aligned}$$

Furthermore, suppose $\min(|a'_2|, |a'_3|, b'_{23}) = |a'_3|$ (again, other subcases can be proven in a similar way), then we can write: $a'_2 = a''_2 + a'_3$ and $b'_{23} = b''_{23} - a'_3$, with $a''_2 \leq 0$ and $b''_{23} \geq 0$. We obtain:

$$\begin{aligned}
&a'_2|B| + a'_3|C| + b'_{12}|A \cap B| + b'_{13}|A \cap C| + b'_{23}|B \cap C| \\
&+ (c + a_1)|X| \\
&\geq a''_2|B| + b'_{12}|A \cap B| + b'_{13}|A \cap C| + b''_{23}|B \cap C| \\
&\quad + \ (c + a_1 + a'_3)|X| \\
&\geq (c + a_1 + a''_2 + a'_3)|X| \\
&\geq 0\,,
\end{aligned}$$

since $c + a_1 + a''_2 + a'_3 = c + a_2 \geq 0$.

(ii') Let us demonstrate that when $\min(|a_1|, |a_2|, |a_3|, b_{12}, b_{13}, b_{23}) = |a_1|$ and $c + a_1 + a_3 \geq 0$, inequality B_3 is not necessary. In that case, we can write $a_2 = a'_2 + a_1$ and $b_{12} = b'_{12} - a_1$, with $a'_2 \leq 0$ and $b'_{12} \geq 0$. We then obtain

$$\begin{aligned}
&a_1|A| + a_2|B| + a_3|C| + b_{12}|A \cap B| + b_{13}|A \cap C| + b_{23}|B \cap C| \\
&+ c|X| \\
&\geq a'_2|B| + a_3|C| + b'_{12}|A \cap B| + b_{13}|A \cap C| + b_{23}|B \cap C| \\
&\quad + \ (c + a_1)|X|\,.
\end{aligned}$$

Furthermore, suppose $\min(|a'_2|, |a_3|, b_{23}) = |a'_2|$ (again, other subcases can be proven in a similar way), then we can write: $a_3 = a'_3 + a'_2$ and $b_{23} = b'_{23} - a'_2$, with $a'_3 \leq 0$ and $b'_{23} \geq 0$. We obtain:

$$\begin{aligned}
&a'_2|B| + a_3|C| + b'_{12}|A \cap B| + b_{13}|A \cap C| + b_{23}|B \cap C| \\
&+ (c + a_1)|X| \\
&\geq a'_3|C| + b'_{12}|A \cap B| + b_{13}|A \cap C| + b'_{23}|B \cap C| \\
&\quad + \ (c + a_1 + a'_2)|X| \\
&\geq (c + a_1 + a'_2 + a'_3)|X| \\
&\geq 0\,,
\end{aligned}$$

since $c + a_1 + a'_2 + a'_3 = c + a_1 + a_3 \geq 0$.

This completes our proof. □

5.5 Conclusions

In this chapter, we have shown that the Bell-type inequalities contribute greatly to the verification of inequalities on fuzzy cardinalities. Moreover, we have formulated two meta-theorems, which state that certain inequalities which are valid in the crisp case stay invariant when fuzzified. These meta-theorems will be extensively used in the next two chapters.

Chapter 6

A parametric family of fuzzy similarity measures

6.1 Introduction

In Chapter 2, we have given an extended overview of commonly used similarity measures for binary vectors denoting the presence or absence of features. However, similarity measurement does not restrict to binary vectors only. Many times, graded feature vectors, i.e. vectors whose elements are scaled to the real unit interval, are to be compared. Therefore, one of the aims of this chapter is to provide the reader with a class of similarity measures for comparing graded feature vectors. We will refer to these similarity measures as fuzzy similarity measures, as will become clear in Section 6.2.

In the literature, fuzzy similarity measures appear in various ways and in several domains. For example, in image processing [15, 63], fuzzy similarity measures are used to compare different images (which are reduced to graded feature vectors). Blanco et al. [11] prove that fuzzy similarity measures can be used for calculating the effectiveness in information retrieval, rather than the traditional measures of recall and precision. In fuzzy modelling [16], a fuzzified version of the Jaccard coefficient (not only based on the basic scalar cardinality of a fuzzy set, but also on the cardinality of a fuzzy set which is characterized by a symmetrical, continuous membership function) can be used to remove redundant fuzzy rules and to minimize the number of fuzzy sets.

It is always possible to derive in an axiomatic setting fuzzy similarity measures from scratch. Nevertheless, a simpler method to construct fuzzy similarity measures is to start from a similarity measure for ordinary sets and to establish a set of fuzzification rules by which the measure is transformed in a consistent manner into a fuzzy similarity measure. In most papers concerned with fuzzy similarity measures, according to a well-known recipe of Zadeh [95], the minimum and maximum operator are used to model pointwise intersection and union of fuzzy sets and the basic scalar cardinality is used to define the cardinality of a fuzzy set. Many papers have already been dedicated to the fuzzification of cardinality-based similarity measures. They are mainly based on the fuzzification of Tversky's contrast model [81].

The fuzzified version of the Jaccard coefficient is by far the most popular fuzzy similarity measure [6, 10, 15, 64], but also fuzzifications of the simple matching coefficient [10], the Dice coefficient [15, 64] and the Ochiai coefficient [63] appear. Furthermore, new families of fuzzy similarity measures based on Tversky's contrast model were given by Tolias et al. [80] and by Santini and Jain [71]. Next to the fuzzification of the Jaccard coefficient and the Dice coefficient, Pappis and Karacapilidis [64] also provide a fuzzification of the following similarity measure for ordinary sets $S(A,B) = 1 - |A \triangle B|$ (which also appeared in Dubois and Prade [29]).

Wang, De Baets and Kerre [84] take a different direction. They provide a class of fuzzy similarity measures based on the work by Bandler and Kohout [2] and inspired on the classical equivalence $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$. Considering different implicators (an implicator is a $[0,1]^2 \to [0,1]$ mapping for which $\mathcal{I}(0,0) = \mathcal{I}(0,1) = \mathcal{I}(1,1) = 1$ and $\mathcal{I}(1,0) = 0$ and whose first (second) partial mappings are decreasing (increasing)), a class of similarity measures $E_{\mathcal{I}}(A,B) = \min(\inf_x \mathcal{I}(A(x),B(x)), \inf_x \mathcal{I}(B(x),A(x)))$ is then obtained. One particular member of this class, the fuzzified version of the above-mentioned similarity measure based on the symmetric difference, is obtained by considering the Łukasiewicz implicator, i.e. $\mathcal{I}(x,y) = \min(1-x+y,1)$.

Bouchon-Meunier et al. [10] propose a classification of measures of comparison enabling to compare fuzzy characteristics of objects, according to their properties and the purpose of their utilization. These measures of comparison are subdivided into measures of dissimilarity and measures of similitude (both are based on Tversky's contrast model). The latter are subsequently divided into measures of satisfiability and measures of resemblance.

This chapter is organized as follows. In Section 6.2, a definition of a fuzzy similarity measure is given and a family of fuzzy similarity measures is proposed, based on the fuzzification of a family of cardinality-based similarity measures for ordinary sets introduced in Chapter 2. In Section 6.3 the $T_{\mathbf{L}}$- and $T_{\mathbf{P}}$-transitive members are identified.

6.2 A parametric family of fuzzy similarity measures

Similarity measurement does not restrict to binary vectors only. Often, the presence or absence of a feature is not clear-cut and is rather a matter of degree. Hence, if instead of binary vectors we have to compare vectors with components scaled to the real unit interval $[0, 1]$ (the higher the number, the more the feature is present), the need arises to generalize the similarity measures described in Chapter 2. In fact, in the same way as binary vectors can be identified with ordinary subsets of a finite universe X, vectors with components in $[0, 1]$ can be identified with fuzzy subsets of X. We define a fuzzy similarity measure as follows:

Definition 6.1 *A fuzzy similarity measure is a reflexive, symmetrical binary fuzzy relation on $\mathcal{F}(X)$.*

Having introduced fuzzification rules for the cardinality of a fuzzy set and translated classical set operations in Chapter 3, we are now able to fuzzify the parametric family of similarity measures (2.4). Therefore, we rewrite expression (2.4) in terms of intersections only,

$$\begin{aligned}&S(A,B)\\&= \frac{x(|A|+|B|-2|A\cap B|)+y|A\cap B|+z(n-|A|-|B|+|A\cap B|)}{x'(|A|+|B|-2|A\cap B|)+y|A\cap B|+z(n-|A|-|B|+|A\cap B|)}\,.\end{aligned}$$

Consider two fuzzy sets A and B in a finite universe $X = \{x_1, \ldots, x_n\}$, then we fuzzify the above expression into

$$S(A,B) = \frac{x(a+b-2u)+yu+z(n-a-b+u)}{x'(a+b-2u)+yu+z(n-a-b+u)}\,, \tag{6.1}$$

with $a = \sum_{i=1}^{n} A(x_i)$, $b = \sum_{i=1}^{n} B(x_i)$ and $u = \sum_{i=1}^{n} Q(A(x_i), B(x_i))$, where Q denotes a commutative quasi-copula. Remark that if we model the intersection of two fuzzy sets by means of a stable, commutative quasi-copula, it doesn't matter whether we first rewrite expression (2.4) in

Similarity measure	$T_{\mathbf{M}}$	$T_{\mathbf{P}}$
Jaccard	$\frac{\sum_{i=1}^{n} \min(A(x_i),B(x_i))}{\sum_{i=1}^{n} A(x_i)+B(x_i)-\min(A(x_i),B(x_i))}$	$\frac{\sum_{i=1}^{n} A(x_i)B(x_i)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-A(x_i)B(x_i)}$
Simple matching	$\frac{\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2\min(A(x_i),B(x_i))}{n}$	$\frac{\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2A(x_i)B(x_i)}{n}$
Dice	$\frac{2\sum_{i=1}^{n} \min(A(x_i),B(x_i))}{\sum_{i=1}^{n} A(x_i)+B(x_i)}$	$\frac{2\sum_{i=1}^{n} A(x_i)B(x_i)}{\sum_{i=1}^{n} A(x_i)+B(x_i)}$
Rogers and Tanimoto	$\frac{\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2\min(A(x_i),B(x_i))}{\sum_{i=1}^{n} 1+A(x_i)+B(x_i)-2\min(A(x_i),B(x_i))}$	$\frac{\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2A(x_i)B(x_i)}{\sum_{i=1}^{n} 1+A(x_i)+B(x_i)-2A(x_i)B(x_i)}$
Sokal and Sneath 1	$\frac{\sum_{i=1}^{n} \min(A(x_i),B(x_i))}{\sum_{i=1}^{n} 2A(x_i)+2B(x_i)-3\min(A(x_i),B(x_i))}$	$\frac{\sum_{i=1}^{n} A(x_i)B(x_i)}{\sum_{i=1}^{n} 2A(x_i)+2B(x_i)-3A(x_i)B(x_i)}$
Sokal and Sneath 2	$\frac{2\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2\min(A(x_i),B(x_i))}{\sum_{i=1}^{n} 2-A(x_i)-B(x_i)+2\min(A(x_i),B(x_i))}$	$\frac{2\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2A(x_i)B(x_i)}{\sum_{i=1}^{n} 2-A(x_i)-B(x_i)+2A(x_i)B(x_i)}$

Table 6.1: Fuzzy similarity measures when using $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$.

Similarity measure	$T_{\mathbf{L}}$
Jaccard	$\frac{\sum_{i=1}^{n} \max(A(x_i)+B(x_i)-1,0)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-\max(A(x_i)+B(x_i)-1,0)}$
Simple matching	$\frac{\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2\max(A(x_i)+B(x_i)-1,0)}{n}$
Dice	$\frac{2\sum_{i=1}^{n} \max(A(x_i)+B(x_i)-1,0)}{\sum_{i=1}^{n} A(x_i)+B(x_i)}$
Rogers and Tanimoto	$\frac{\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2\max(A(x_i)+B(x_i)-1,0)}{\sum_{i=1}^{n} 1+A(x_i)+B(x_i)-2\max(A(x_i)+B(x_i)-1,0)}$
Sokal and Sneath 1	$\frac{\sum_{i=1}^{n} \max(A(x_i)+B(x_i)-1,0)}{\sum_{i=1}^{n} 2A(x_i)+2B(x_i)-3\max(A(x_i)+B(x_i)-1,0)}$
Sokal and Sneath 2	$\frac{2\sum_{i=1}^{n} 1-A(x_i)-B(x_i)+2\max(A(x_i)+B(x_i)-1,0)}{\sum_{i=1}^{n} 2-A(x_i)-B(x_i)+2\max(A(x_i)+B(x_i)-1,0)}$

Table 6.1: Fuzzy similarity measures when using $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ (continued).

terms of intersections only and then fuzzify this new expression, or whether we fuzzify expression (2.4) directly. In Table 6.1, we summarize the fuzzy similarity measures (that are a member of family (6.1) and whose crisp counterparts can be found in Table 2.6) for some commonly used quasi-copulas, $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$.

Again, remark that in order to guarantee that $S(A,B) \in [0,1]$, we need to impose the following restriction: $0 \leq x \leq x'$. Analogously to the parametric family of similarity measures for ordinary sets, the case $x = x'$ leads to trivial measures taking value 1 only, and therefore we consider from here on $0 \leq x < x'$.

6.3 T-transitive members

It is highly desirable that T-transitivity is preserved along the fuzzification process from crisp to fuzzy similarity measures. First, we identify the $T_{\mathbf{L}}$-transitive members of family (6.1) to continue with the $T_{\mathbf{P}}$-transitive members. These results are in line with the ones obtained for the family of crisp similarity measures.

6.3.1 Łukasiewicz-transitive members

Let us recall that a fuzzy similarity measure is $T_{\mathbf{L}}$-transitive if the following inequality is fulfilled:

$$1 - S(A,B) - S(B,C) + S(A,C) \geq 0\,. \tag{6.2}$$

Theorem 6.1 *The $T_{\mathbf{L}}$-transitive members of family (6.1) of fuzzy similarity measures are for any commutative quasi-copula that satisfies I_3^3 characterized by:*

$$x' \geq \max(y,z)\,. \tag{6.3}$$

Proof. In Theorem 2.1 we proved that inequality (6.2) holds for all n and all ordinary sets A, B and C if and only if the parameters satisfy conditions (6.3). For fuzzy sets A, B and C, the left-hand side of inequality (6.2) is a homogeneous function of $|A|$, $|B|$, $|C|$, $|A\cap B|$, $|A\cap C|$, $|B\cap C|$ and $|X|$. When inequality I_3^3 is fulfilled (inequalities I_2^1 and I_3^2 are satisfied for any quasi-copula), we can use Theorem 5.3 to conclude that inequality (6.2) also holds for all fuzzy sets A, B and C under the same parameter conditions (6.3). □

Theorem 6.2 *The $T_{\mathbf{L}}$-transitive members of family (6.1) of fuzzy similarity measures with $z = 0$ are for any commutative quasi-copula characterized by parameter conditions (6.3).*

Proof. When $z = 0$, the homogeneous function on the left-hand side of (6.2) is independent of $|X|$ and using Theorem 5.4 we can conclude that inequality (6.2) also holds for all fuzzy sets A, B and C under the same parameter conditions (6.3). □

Using the latter theorem, we can conclude that the fuzzified version of the Jaccard coefficient as well as the fuzzified version of the first coefficient of Sokal and Sneath are $T_{\mathbf{L}}$-transitive, since both are members of family (6.1) with $z = 0$.

However, with Theorem 6.1 we are not able to verify the $T_{\mathbf{L}}$-transitivity of other fuzzy similarity measures when we model the intersection of two fuzzy sets by a commutative quasi-copula which does not satisfy Bell-type inequality I_3^3. To solve this problem, we provide a third theorem. It requires a direct proof and has nothing to do with the meta-theorems discussed above. First, we prove the following lemmata:

Lemma 6.1 *Let $s_1, s_2, s_3 \in \mathbb{R}$ with $s_1 > 0$, $s_2 > 0$ and $s_3 > 0$ such that*

$$\frac{1}{s_1} + \frac{1}{s_2} - \frac{1}{s_3} \geq 0\,,$$

then also the following holds

$$\frac{1}{s_1 + t} + \frac{1}{s_2 + t} - \frac{1}{s_3 + t} \geq 0\,,$$

for any $t \geq 0$.

Proof. Since the denominators of the left part of the latter inequality are always positive, we have to verify that the following inequality is fulfilled:

$$(s_1 + t)(s_3 + t) + (s_2 + t)(s_3 + t) - (s_1 + t)(s_2 + t) \geq 0 \tag{6.4}$$

for any $t \geq 0$. Inequality (6.4) can be rewritten as

$$t^2 + 2s_3t + s_3(s_1 + s_2) - s_1s_2 \geq 0\,.$$

Since $\frac{1}{s_1} + \frac{1}{s_2} - \frac{1}{s_3} \geq 0$, or equivalently $s_3(s_1 + s_2) - s_1s_2 \geq 0$, and $s_3 > 0$, inequality (6.4) is fulfilled for any $t \geq 0$. □

Lemma 6.2 *Let f be a quadratic function in x, defined as $f(x) = \alpha x^2 + \beta x + \gamma$, with α, β and γ real coefficients such that $\gamma = f(0) \geq 0$. If there exists a $y > 0$ such that $f(y) \geq 0$ and $2\gamma + \beta y \geq 0$, then $f(x) \geq 0$ for any $x \in [0, y]$.*

Proof. Since $f(0) \geq 0$ and $f(y) \geq 0$ for $y > 0$, the quadratic function f could only change sign in $[0, y]$ if $f'(0) = \beta < 0$, $f'(y) = 2\alpha y + \beta > 0$ and $f''(0) = \alpha > 0$. Suppose $\alpha > 0$ and $0 < -\beta < 2\alpha y$. Then, the discriminant of the quadratic function f is equal to

$$\triangle = \beta^2 - 4\alpha\gamma < -2\alpha\beta y - 4\alpha\gamma = 2\alpha(-\beta y - 2\gamma)\,.$$

Since $2\gamma + \beta y \geq 0$, we can conclude that $\triangle \leq 0$ and therefore the function f is always positive in $[0, y]$. □

In the next two lemmata, we prove that some inequalities on fuzzy set cardinalities are always fulfilled (these inequalities will show up in the proof of Theorem 6.3). We introduce the following notations:

$$\begin{aligned}
a &= \sum_{i=1}^{n} A(x_i)\,, & u &= \sum_{i=1}^{n} Q(A(x_i), B(x_i))\,,\\
b &= \sum_{i=1}^{n} B(x_i)\,, & v &= \sum_{i=1}^{n} Q(A(x_i), C(x_i))\,,\\
c &= \sum_{i=1}^{n} C(x_i)\,, & w &= \sum_{i=1}^{n} Q(B(x_i), C(x_i))\,,
\end{aligned}$$

where Q denotes a commutative quasi-copula.

Lemma 6.3 *Consider the notations introduced above, then the inequality*

$$\begin{aligned}
&(a + c)(a + b - 2u)(b + c - 2w) + uv(b + c - 2w)\\
&- uw(a + c - 2v) + vw(a + b - 2u) \geq 0
\end{aligned} \tag{6.5}$$

is always fulfilled.

Proof. Let us consider the expression

$$E = c(b - u)(b + c - 2w) + uv(c - w) - uw(c - v) + vw(b - u)\,,$$

which can also be written as

$$E = c(b - u)(b + c - 2w) + uc(v - w) + vw(b - u)\,. \tag{6.6}$$

Since inequality I_3^2 is fulfilled for any commutative quasi-copula, the following inequality is satisfied: $v-u-w+b \geq 0$. Additionally, making use of the fact that $c \geq 0$ and $u \geq 0$, we obtain from (6.6) that

$$\begin{aligned} E &\geq c(b-u)(b+c-2w) + uc(u-b) + vw(b-u) \\ &= [c(b+c-2w-u) + vw](b-u)\,. \end{aligned}$$

Since $c \geq 0$ and $b-u \geq 0$, and since inequality I_3^2 also yields that $b-u \geq w-v$, it follows that:

$$\begin{aligned} E &\geq [c(c-2w+w-v) + vw](b-u) \\ &= [c(c-w) - v(c-w)](b-u) = (c-v)(c-w)(b-u)\,. \end{aligned}$$

Hence, $E \geq 0$, and adding to E the positive quantity $c(a-u)(b+c-2w)$ enforces the inequality, so that:

$$c(a+b-2u)(b+c-2w) + uv(c-w) - uw(c-v) + vw(b-u) \geq 0\,. \tag{6.7}$$

This inequality holds for any A, B and C and corresponding a, b and c. Therefore, we may change the role of A and C, which implies that we can simultaneously change a into c, c into a, u into w and w into u to obtain that:

$$a(a+b-2u)(b+c-2w) + uv(b-w) - uw(a-v) + vw(a-u) \geq 0\,. \tag{6.8}$$

The side by side addition of inequalities (6.7) and (6.8) results in inequality (6.5). □

Lemma 6.4 *Consider the notations introduced above, then the inequality*

$$\begin{aligned} &4n^2(v-u-w+b) - n[(a+b-2u)(v+w) \\ &+ (a+c-2v)(u+w) - (b+c-2w)(u+v)] \geq 0 \end{aligned} \tag{6.9}$$

is always fulfilled.

Proof. Since $n \geq 0$, it suffices to show that the following inequality is fulfilled:

$$\begin{aligned} &4n(v-u-w+b) - (a+b-2u)(v+w) \\ &- (a+c-2v)(u+w) - (b+c-2w)(u+v) \geq 0\,. \end{aligned}$$

Since the latter inequality should hold for any fuzzy sets A, B and C, we can substitute the fuzzy sets by their complements. Then the following inequality should be fulfilled:

$$\begin{aligned}
&4n(b-u-w+v)-(a+b-2u)(n-a-c+v+n-b-c+w)\\
&-(a+c-2v)(n-a-b+u+n-b-c+w)\\
&+(b+c-2w)(n-a-b+u+n-a-c+v)\geq 0\,.
\end{aligned}\tag{6.10}$$

The left-hand side of inequality (6.10) can be rewritten as follows:

$$\begin{aligned}
&4n(b-u-w+v)-(a+b-2u)(n-a-c+v+n-b-c+w)\\
&-(a+c-2v)(n-a-b+u+n-b-c+w)\\
&+(b+c-2w)(n-a-b+u+n-a-c+v)\\
=\;&(a+b-2u)(a+b+2c-v-w)+(a+c-2v)(a+2b+c-u-w)\\
&-(b+c-2w)(2a+b+c-u-v)\\
=\;&(a+b-2u)(c-v)+(a+c-2v)(b-u)\\
&+(2a+b+c)(a-u-v+w)\,.
\end{aligned}$$

Since inequality I_3^2 is fulfilled, also inequality (6.10) will hold. □

Theorem 6.3 *The $T_{\mathbf{L}}$-transitive members of family (6.1) of fuzzy similarity measures are for any commutative quasi-copula characterized by parameter conditions (6.3).*

Proof. In order to identify the conditions on the parameters x', y and z in (6.1), we have to verify when the following inequality is fulfilled:

$$\begin{aligned}
(x'-x)\Bigg(-&\frac{a+c-2v}{x'(a+c-2v)+y\,v+z(n-a-c+v)}\\
+&\frac{a+b-2u}{x'(a+b-2u)+y\,u+z(n-a-b+u)}\\
+&\frac{b+c-2w}{x'(b+c-2w)+y\,w+z(n-b-c+w)}\Bigg)\geq 0\,.
\end{aligned}\tag{6.11}$$

Again, we can omit the factor $x'-x$, since $x'>x$.
Case $z=0$**.**
Substituting $z=0$ in inequality (6.11) and converting the fractions of the left-hand side of this inequality so that they have a common denominator, it is sufficient to study the numerator, since the denominator is

always positive. Then, the following inequality should be verified:

$$\begin{aligned}K(x',y) = {} & x'^2(a+b-2u)(a+c-2v)(b+c-2w)\\ & +2x'y(a+b-2u)(b+c-2w)v\\ & +y^2(uv(b+c-2w)-uw(a+c-2v)+vw(a+b-2u))\\ \geq {} & 0\,.\end{aligned}$$

Since inequality (6.5) is satisfied (which was proven in Lemma 6.4), we can conclude that

$$\begin{aligned}& uv(b+c-2w)-uw(a+c-2v)+vw(a+b-2u)\\ \geq {} & -(a+c)(a+b-2u)(b+c-2w)\end{aligned}$$

and therefore the following will hold:

$$\begin{aligned}K(x',y) \geq {} & (x'^2-y^2)(a+b-2u)(a+c-2v)(b+c-2w)\\ & +2y(x'-y)(a+b-2u)(b+c-2w)v\,.\end{aligned}$$

It is obvious that $K(x',y) \geq 0$ if $x' \geq y$. Therefore, the condition

$$x' \geq y \,\wedge\, z = 0$$

is a sufficient condition for inequality (6.11) to hold.

Case $y = 0$**.**
Since $a+b-2u = (n-a)+(n-b)-2(n-a-b+u)$ (and similar equalities for v and w), this case is completely analogous to the previous one provided that the roles of y and z are interchanged and that one turns into complements. Therefore, the condition

$$x' \geq z \,\wedge\, y = 0$$

is a sufficient condition for inequality (6.11) to hold.

Case $y \neq 0,\ z \neq 0$**.**
Since changing the fuzzy sets into their complements leads to changing the role of y and z, we can suppose $y \leq z$. We will study the following

expression:

$$\begin{aligned}
L(x', y, z) \;=\; & \frac{a+b-2u}{x'(a+b-2u)+y\,u+z(n-a-b+u)} \\
& -\frac{a+c-2v}{x'(a+c-2v)+y\,v+z(n-a-c+v)} \\
& +\frac{b+c-2w}{x'(b+c-2w)+y\,w+z(n-b-c+w)} \\
\;=\; & \frac{1}{x'+(y\,u+z(n-a-b+u))/(a+b-2u)} \\
& -\frac{1}{x'+(y\,v+z(n-a-c+v))/(a+c-2v)} \\
& +\frac{1}{x'+(y\,w+z(n-b-c+w))/(b+c-2w)}\,.
\end{aligned}$$

Due to Lemma 6.1 it is sufficient to prove that $L(z, y, z) \geq 0$ in order that $L(x', y, z) \geq 0$ for any $x' \geq z$. We obtain

$$\begin{aligned}
L(z, y, z) \;=\; & \frac{a+b-2u}{z(n-u)+y\,u} - \frac{a+c-2v}{z(n-v)+y\,v} + \frac{b+c-2w}{z(n-w)+y\,w} \\
\;=\; & \frac{1}{z}\left\{\frac{a+b-2u}{n-\mu\,u} - \frac{a+c-2v}{n-\mu\,v} + \frac{b+c-2w}{n-\mu\,w}\right\},
\end{aligned}$$

with $\mu = \frac{y-z}{z}$, taking values between 0 (i.e. $y = z$) and 1 (i.e. $y = 0$). Now, the problem reduces to proving that $L(z, y, z) \geq 0$ for any $\mu \in [0, 1]$. Converting the fractions of the latter expression for $L(z, y, z)$ so that they have a common denominator and omitting this denominator since it is always positive, we have to verify that

$$\begin{aligned}
M(\mu) \;=\; & (n-\mu\,v)[(a+b-2u)(n-\mu\,w)+(b+c-2w)(n-\mu\,u)] \\
& -(a+c-2v)(n-\mu\,u)(n-\mu\,w) \\
\;=\; & \mu^2[(a+b-2u)vw-(a+c-2v)uw+(b+c-2w)uv] \\
& -\mu\,n[(a+b-2u)(v+w)-(a+c-2v)(u+w) \\
& +(b+c-2w)(u+v)]+2n^2(v-u-w+b)
\end{aligned}$$

is always positive for any $\mu \in [0, 1]$. Since inequality I_3^2 is fulfilled, also $M(0) = 2n^2(b-u-w+v) \geq 0$ will hold. Since $\mu = 1$ corresponds to $y = 0$, we can also conclude that $M(1) \geq 0$. This was proven in the previous case. To apply Lemma 6.2, we still have to verify that the following inequality will hold:

$$\begin{aligned}
& 4n^2(v-u-w+b) - n[(a+b-2u)(v+w) \\
& +\;(a+c-2v)(u+w)-(b+c-2w)(u+v)] \geq 0\,.
\end{aligned}$$

It was proven in Lemma 6.4 that the latter inequality is always satisfied. Due to Lemma 6.2, $M(\mu) \geq 0$ for any $\mu \in [0,1]$. We can conclude that

$$x' > x \,\wedge\, x' \geq \max(y, z)\,,$$

are sufficient conditions such that the members of family (6.1) of fuzzy similarity measure are $T_{\mathbf{L}}$-transitive.

The proof that these conditions are also necessary conditions is already given in Theorem 2.1. □

Corollary 6.1 *The $T_{\mathbf{L}}$-transitive members of family (6.1) of fuzzy similarity measures are for any Frank t-norm characterized by parameter conditions (6.3).*

This corollary immediately follows from Theorem 6.3. It only follows from Theorem 6.1 for $\lambda \leq 9+4\sqrt{5}$ or from Theorem 6.2 when $z = 0$ (this is the case for the Jaccard coefficient and the first coefficient of Sneath and Sokal).

6.3.2 Product-transitive members

The same reasoning can be made for identifying the parameters of the family of fuzzy similarity measures such that the members of this family are $T_{\mathbf{P}}$-transitive. Let us recall that a fuzzy similarity measure is $T_{\mathbf{P}}$-transitive if the following inequality is fulfilled:

$$S(A,C) - S(A,B) \cdot S(B,C) \geq 0\,. \tag{6.12}$$

Theorem 6.4 *The $T_{\mathbf{P}}$-transitive members of family (6.1) of fuzzy similarity measures are for any commutative quasi-copula that satisfies I_3^3 characterized by:*

$$x\,x' \geq \max(y^2,\, z^2)\,. \tag{6.13}$$

Proof. In Theorem 2.2 we proved that inequality (6.12) holds for all n and all ordinary sets A, B and C if and only if the parameters satisfy (6.13). For fuzzy sets A, B and C, the left-hand side of (6.12) is a homogeneous function of $|A|$, $|B|$, $|C|$, $|A \cap B|$, $|A \cap C|$, $|B \cap C|$ and $|X|$. When inequality I_3^3 is fulfilled (inequalities I_2^1 and I_3^2 are satisfied for any quasi-copula), we can use Theorem 5.3 to conclude that inequality (6.12) also holds for all fuzzy sets A, B and C under the same parameter conditions (6.13). □

Theorem 6.5 *The $T_{\mathbf{P}}$-transitive members of family (6.1) of fuzzy similarity measures with $z = 0$ are for any commutative quasi-copula characterized by parameter conditions (6.13).*

Proof. When $z = 0$, the homogeneous function on the left-hand side of inequality (6.12) is independent of $|X|$ and using Theorem 5.4 we can conclude that inequality (6.12) also holds for all fuzzy sets A, B and C under the same parameter conditions (6.13). □

Again, we are not able to verify the $T_{\mathbf{P}}$-transitivity of fuzzy similarity measures in which the intersection of two fuzzy sets is modelled by a commutative quasi-copula which does not satisfy inequality I_3^3. Therefore, we provide a direct algebraic proof to overcome this problem.

Theorem 6.6 *The $T_{\mathbf{P}}$-transitive members of family (6.1) of fuzzy similarity measures are for any commutative quasi-copula characterized by parameter conditions (6.13).*

Proof. In order to identify the conditions on the parameters x, x', y and z in (6.1), we have to verify when the following inequality is fulfilled:

$$(x'-x)\left(\frac{a+b-2u}{N_1} - \frac{a+c-2v}{N_2} + \frac{b+c-2w}{N_3} \right. \\ \left. - \frac{(x'-x)(a+b-2u)(b+c-2w)}{N_1\,N_3}\right) \geq 0\,. \tag{6.14}$$

with,

$$\begin{aligned} N_1 &= x'(a+b-2u) + y\,u + z(n-a-b+u)\,,\\ N_2 &= x'(a+c-2v) + y\,v + z(n-a-c+v)\,,\\ N_3 &= x'(b+c-2w) + y\,w + z(n-b-c+w)\,. \end{aligned}$$

Again, we can omit the factor $x'-x$, since $x' > x$. After converting the fractions of the left part of inequality (6.14) so that they have a common denominator and omitting this denominator since it is always positive, we divide the expression by x'. In addition suppose $\overline{x} = x/x'$, $\overline{y} = y/x'$ and $\overline{z} = z/x'$ such that $0 \leq \overline{x} < 1$, $0 \leq \overline{y} \leq 1$ and $0 \leq \overline{z} \leq 1$, then we have to identify the conditions on the parameters $\overline{x}, \overline{y}$ and $\overline{z}$ such that the following expression is greater or equal to zero:

$$\begin{aligned}K(\overline{x},\overline{y},\overline{z}) = {} & \overline{x}\,(a+b-2u)(b+c-2w)(a+c-2v)\\
& + \overline{y}^2[(a+b-2u)vw-(a+c-2v)uw+(b+c-2w)uv]\\
& + \overline{z}^2[(a+b-2u)(n-a-c+v)(n-b-c+w)\\
& \quad -(a+c-2v)(n-a-b+u)(n-b-c+w)\\
& \quad +(b+c-2w)(n-a-b+u)(n-a-c+v)]\\
& + \overline{y}(1+\overline{x})(a+b-2u)(b+c-2w)v\\
& + \overline{z}(1+\overline{x})(a+b-2u)(b+c-2w)(n-a-c+v)\\
& + \overline{yz}[(a+b-2u)(v(n-c-b+w)+w(n-a-c+v))\\
& \quad +(b+c-2w)(v(n-a-b+u)+u(n-a-c+v))\\
& \quad -(a+c-2v)(w(n-a-b+u)+u(n-b-c+w))]\,.\end{aligned}$$

Remark that $K(\overline{x},\overline{y},\overline{z})$ is a non-decreasing function of $\overline{x}$. Therefore, if we can prove that $K(\overline{x}_0,\overline{y},\overline{z}) \geq 0$ for $\overline{x}_0 = \max(\overline{y}^2,\overline{z}^2)$, then $K(\overline{x},\overline{y},\overline{z}) \geq 0$ for any $\overline{x} \in [\overline{x}_0, 1[$.

When we replace the fuzzy sets in expression (6.14) by their complements, then only the parameters y and z are changed and therefore also $\overline{y}$ and $\overline{z}$. Therefore, we can suppose that $z \leq y$ and also $\overline{z} \leq \overline{y}$. Let $\overline{x}_0 = \overline{y}^2$, then

$$\begin{aligned}K(\overline{y}^2,\overline{y},\overline{z}) = {} & \overline{y}^2\,(a+b-2u)(b+c-2w)(a+c-2v)\\
& + \overline{y}^2\,[(a+b-2u)vw-(a+c-2v)uw+(b+c-2w)uv]\\
& + \overline{z}^2\,[(a+b-2u)(n-a-c+v)(n-b-c+w)\\
& \quad -(a+c-2v)(n-a-b+u)(n-b-c+w)\\
& \quad +(b+c-2w)(n-a-b+u)(n-a-c+v)]\\
& + \overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)v\\
& + \overline{z}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)(n-a-c+v)\\
& + \overline{yz}\,[(a+b-2u)(v(n-c-b+w)+w(n-a-c+v))\\
& \quad +(b+c-2w)(v(n-a-b+u)+u(n-a-c+v))\\
& \quad -(a+c-2v)(w(n-a-b+u)+u(n-b-c+w))]\,.\end{aligned}$$

This expression can be considered as a quadratic form $\alpha\overline{z}^2+\beta\overline{z}+\gamma$ in $\overline{z}$, with $\overline{z} \in [0,\overline{y}]$. We will use Lemma 6.2 to prove that $K(\overline{y}^2,\overline{y},\overline{z}) \geq 0$ as follows: first, we prove that $K(\overline{y}^2,\overline{y},0) = \gamma \geq 0$, next we prove that $K(\overline{y}^2,\overline{y},\overline{y}) = \alpha+\beta+\gamma \geq 0$ and finally we prove that $2\gamma+\beta\overline{y} \geq 0$.

Substituting $\overline{z} = 0$, we obtain:

$$\begin{aligned} K(\overline{y}^2, \overline{y}, 0) \,=\, & \overline{y}^2\,(a+b-2u)(b+c-2w)(a+c-2v) \\ & + \overline{y}^2\,[(a+b-2u)vw - (a+c-2v)uw + (b+c-2w)uv] \\ & + \overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)v\,. \end{aligned}$$

From inequality (6.5) it follows that

$$\begin{aligned} K(\overline{y}^2, \overline{y}, 0) \,\geq\, & -2\overline{y}^2\,(a+b-2u)(b+c-2w)v \\ & + \overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)v \\ =\, & \overline{y}(1-\overline{y})^2\,(a+b-2u)(b+c-2w)v\,. \end{aligned}$$

Since $0 \leq \overline{y} \leq 1$, $a+b-2u \geq 0$ and $b+c-2w \geq 0$, we can conclude that $K(\overline{y}^2, \overline{y}, 0) \geq 0$. Substituting $\overline{z} = \overline{y}$, we obtain the following:

$$\begin{aligned} K(\overline{y}^2, \overline{y}, \overline{y}) = & \,\overline{y}^2\,\,((a+b-2u)(b+c-2w)(a+c-2v) \\ & + (a+b-2u)vw - (a+c-2v)uw + (b+c-2w)uv \\ & + (a+b-2u)(n-a-c+v)(n-b-c+w) \\ & - (a+c-2v)(n-a-b+u)(n-b-c+w) \\ & + (b+c-2w)(n-a-b+u)(n-a-c+v) \\ & + (a+b-2u)(v(n-c-b+w) + w(n-a-c+v)) \\ & + (b+c-2w)(v(n-a-b+u) + u(n-a-c+v)) \\ & -\,(a+c-2v)(w(n-a-b+u) + u(n-b-c+w))) \\ & + \overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)(n-a-c+v) \\ = & \,\overline{y}^2\,\,((a+b-2u)(b+c-2w)(a+c-2v) \\ & + (a+b-2u)(n-a-c+2v)(n-b-c+w) \\ & - (a+c-2v)(n-a-b+2u)(n-b-c+w) \\ & +\,(b+c-2w)(n-a-b+2u)(n-a-c+v)) \\ & + \overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)(n-a-c+v)\,. \end{aligned}$$

Furthermore, we can write $K(\overline{y}^2,\overline{y},\overline{y})$ as follows:

$$\begin{aligned}
K(\overline{y}^2,\overline{y},\overline{y}) &= \overline{y}^2\ \big(2(a+b-2u)(b+c-2w)(a+c-2v)\\
&\quad -2n(a+b-2u)(b+c-2w)+2n^2(v-u-w+b)\big)\\
&\quad +\overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)(n-a-c+v)\\
&= 2\overline{y}^2n^2(v-u-w+b)\\
&\quad -2\overline{y}^2(a+b-2u)(b+c-2w)(n-a-c+v)\\
&\quad +\overline{y}(1+\overline{y}^2)(a+b-2u)(b+c-2w)(n-a-c+v)\\
&= 2\overline{y}^2n^2(v-u-w+b)\\
&\quad +\overline{y}(1-\overline{y})^2(a+b-2u)(b+c-2w)(n-a-c+v)\,.
\end{aligned}$$

Since Bell inequalities I_2^1 and I_3^2 are fulfilled for any commutative quasi-copula and since $0\le\overline{y}\le 1$, we can conclude that $K(\overline{y}^2,\overline{y},\overline{y})\ge 0$.

We can conclude that $K(\overline{y}^2,\overline{y},\overline{z})\ \ge\ 0$ for any $\overline{z}\in[0,\overline{y}]$ using Lemma 6.2 if we can prove that the following expression

$$\begin{aligned}
L(\overline{y}) &= 2\overline{y}^2\,((a+b-2u)(b+c-2w)(a+c-2v)\\
&\quad +(a+b-2u)vw-(a+c-2v)uw+(b+c-2w)uv)\\
&\quad +2\overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)v\\
&\quad +\overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)(n-a-c+v)\\
&\quad +\overline{y}^2\,((a+b-2u)(v(n-c-b+w)+w(n-a-c+v))\\
&\qquad +(b+c-2w)(v(n-a-b+u)+u(n-a-c+v))\\
&\qquad -(a+c-2v)(w(n-a-b+u)+u(n-b-c+w)))\,.
\end{aligned}$$

or equivalently,

$$\begin{aligned}
L(\overline{y}) &= \overline{y}^2\,(2(a+b-2u)(b+c-2w)(a+c-2v)\\
&\quad +n\,[(a+b-2u)(v+w)-(a+c-2v)(u+w)\\
&\quad +(b+c-2w)(u+v)]-2(a+b-2u)(b+c-2w)v)\\
&\quad +\overline{y}(1+\overline{y}^2)\,(a+b-2u)(b+c-2w)(n-a-c+3v)\\
&= \overline{y}\,\{\overline{y}^2\,(a+b-2u)(b+c-2w)(n-a-c+3v)\\
&\quad +\overline{y}\,(2(a+b-2u)(b+c-2w)(a+c-3v)\\
&\quad +n\,((a+b-2u)(v+w)-(a+c-2v)(u+w)\\
&\quad +(b+c-2w)(u+v)))\\
&\quad +(a+b-2u)(b+c-2w)(n-a-c+3v)\}
\end{aligned}$$

is greater than or equal to zero. In what follows, we discuss the expression $L^*(\overline{y}) = L(\overline{y})/\overline{y}$, given by

$$\begin{aligned}L^*(\overline{y}) = \overline{y}^2\,&(a+b-2u)(b+c-2w)(n-a-c+3v)\\ &+\overline{y}\,[2(a+b-2u)(b+c-2w)(a+c-3v)\\ &+n\,[(a+b-2u)(v+w)-(a+c-2v)(u+w)\\ &+(b+c-2w)(u+v)]]\\ &+(a+b-2u)(b+c-2w)(n-a-c+3v)\,.\end{aligned}$$

This expression can be considered as a quadratic form in $\overline{y}$, with $\overline{y} \in [0,1]$. Again, we will use Lemma 6.2 to prove that $L^*(\overline{y}) \geq 0$. Note that $n-a-c+3v \geq 0$, such that $L^*(0) \geq 0$. On the other hand it also holds that

$$\begin{aligned}L^*(1) &= n\,[(a+b-2u)(v+w)-(a+c-2v)(u+w)\\ &\quad+(b+c-2w)(u+v)+2(a+b-2u)(b+c-2w)]\\ &\geq 0\,.\end{aligned}$$

This can easily be verified by dividing $L^*(1)$ by n and substituting a, b, c, u, v and w in $L^*(1)/n$ by the expressions which hold in the crisp case. In the crisp case, $L^*(1)/n \geq 0$ is always satisfied and therefore Theorem 5.4 can be used to conclude that $L^*(1)/n \geq 0$ also holds for fuzzy cardinalities. In order to prove that $L^*(\overline{y}) \geq 0$ for all $\overline{y} \in [0,1]$ it is sufficient, due to Lemma 6.2, to study the expression

$$\begin{aligned}&\ 2(a+b-2u)(b+c-2w)(n-a-c+3v)\\ &+\ 2(a+b-2u)(b+c-2w)(a+c-3v)\\ &+\ n\,[(a+b-2u)(v+w)-(a+c-2v)(u+w)\\ &+\ (b+c-2w)(u+v)]\,,\end{aligned}$$

which is also equal to

$$\begin{aligned}&n\,[2(a+b-2u)(b+c-2w)+(a+b-2u)(v+w)\\ -\ &(a+c-2v)(u+w)+(b+c-2w)(u+v)]\,.\end{aligned}$$

The latter expression is the same as $L^*(1)$ and therefore it is always greater than or equal to zero.

We can conclude that $L^*(\overline{y}) \geq 0$, and therefore also $L(\overline{y}) \geq 0$ for all $\overline{y} \in [0,1]$. From this, it also follows that $K(\overline{y}^2, \overline{y}, \overline{z}) \geq 0$ for all $\overline{z} \in [0, \overline{y}]$ and $\overline{y} \in [0,1]$. We can conclude that the conditions

$$\overline{x} \geq \max(\overline{y}^2,\ \overline{z}^2)\,,$$

or equivalently,

$$x\,x' \geq \max(y^2,\, z^2)\,,$$

are sufficient conditions such that the members of the family (6.1) of fuzzy similarity measure are $T_{\mathbf{P}}$-transitive.

The proof that these conditions are also necessary conditions is already given in Theorem 2.2. □

Corollary 6.2 *The $T_{\mathbf{P}}$-transitive members of family (6.1) of fuzzy similarity measures are for any Frank t-norm characterized by parameter conditions (6.13).*

Again, this corollary immediately follows from Theorem 6.6. It only follows from Theorem 6.4 for $\lambda \leq 9 + 4\sqrt{5}$ or from Theorem 6.5 when $z = 0$.

6.4 Conclusions

In this chapter, we have fuzzified the parametric family of similarity measures for ordinary sets, which was introduced in Chapter 2, using a commutative quasi-copula to model the intersection of two fuzzy sets. The main result of this chapter is that transitivity, and hence also the corresponding dual metric interpretation, is preserved along this fuzzification process.

Chapter 7

A parametric family of fuzzy inclusion measures

7.1 Introduction

Traditionally, fuzzy set inclusion is defined according to Zadeh's original proposal [95]. For two fuzzy sets A and B, Zadeh defined

$$A \subseteq B \Leftrightarrow (\forall x \in X)(A(x) \leq B(x)) \, ,$$

for any $x \in X$. However, in this sense A is either utterly or not at all a subset of B, whereas one should expect, to be consistent with the spirit of fuzzy set theory, that A is a subset of B to some degree. This observation has led to the introduction of several inclusion measures for fuzzy sets in the literature.

Mainly, two approaches to introduce inclusion measures can be distinguished:

(i) constructing new measures from fuzzy implicators and

(ii) providing an axiomatic approach.

The first method to define a fuzzy inclusion measure consists in a direct fuzzification of the following crisp definition of set inclusion:

$$A \subseteq B \Leftrightarrow (\forall x \in X)(x \in A \Rightarrow x \in B) \, .$$

Using an implicator $\mathcal{I}$ (recall that an implicator is a $[0,1]^2 \to [0,1]$ mapping for which $\mathcal{I}(0,0) = \mathcal{I}(0,1) = \mathcal{I}(1,1) = 1$ and $\mathcal{I}(1,0) = 0$ and whose first (second) partial mappings are decreasing (increasing)), the degree of inclusion of a fuzzy set A in a fuzzy set B is defined as follows:

$$I(A,B) = \inf_{x \in X} \mathcal{I}(A(x), B(x))\,.$$

Bandler and Kohout [2] follow this approach and define several inclusion measures starting from six different implicators. Also Willmott [89] follows this approach and added two inclusion measures to the list of Bandler and Kohout (by using two additional implicators). In [90], Wilmott concluded that all fuzzy inclusion measures defined in [2] and [89] are $T_{\mathbf{L}}$-transitive. Moreover, he proposed a $T_{\mathbf{L}}$-transitive measure based on the cardinality of the sets involved. Later on, De Baets, De Meyer and Naessens [21] proved that this specific inclusion measure is also $T_{\mathbf{P}}$-transitive. Furthermore, Bodenhofer [7] proved that an inclusion measure based on the residual implicator $I_T(x,y) = \sup\{t \in [0,1] \mid T(x,t) \leq y\}$ (with T an arbitrary t-norm) is always T-transitive.

Sinha and Dougherty [76] follow the second approach and postulate a collection of axioms for fuzzy set inclusion in terms of an indicator, also called 'inclusion grade'. Independently of Sinha and Dougherty, Kitainik [51] also developed an axiomatic approach to deal with inclusion. According to Kitainik, a fuzzy inclusion should fulfill four properties (contrapositivity, distributivity, symmetry and heritage).

A combination of these two approaches can also be found. Burillo et al. [13] provide a characterization of a family of inclusion grade operators, based on implicators (a form of generalized Łukasiewicz operators L, defined by $L(x,y) = \min(1, \lambda(x) + \mu(y))$, with λ and μ functions from $[0,1]$ to $[0,1]$ and $\lambda(0) = \mu(1) = 1$ and $\lambda(1) = \mu(0) = 0$). Young [93] presented a set of axioms for fuzzy subsethood that allows one to connect with fuzzy entropy. The author offers these axioms as alternatives to the ones of Sinha and Dougherty. In [33], Fan et al. comment on the subsethood measure defined by Young and give some new definitions of a subsethood measure. They define a subsethood measure starting from a set-theoretic approach and also provide a construction from fuzzy implicators.

Remark that the inclusion measure which was given by Kuncheva is a restriction to the crisp case of the fuzzy inclusion measure investigated by Kosko [55] and Fan et al. [33], and defined by $I(A,B) = M(A \cap B)/M(A)$, where $M(A)$ denotes the cardinality of a fuzzy set A.

Bouchon-Meunier, Rifqi and Bothorel [10] propose a family of inclusion measures as a refinement of a family of measures of similitude. Kosko's inclusion measure is a member of the family defined by Bouchon-Meunier et al.

In this thesis, the starting point is the parametric family of inclusion measures for ordinary sets, which was introduced in Chapter 2.

This chapter is organized as follows. In Section 7.2, we define a fuzzy inclusion measure and propose a family of fuzzy inclusion measures, based on the fuzzification of a family of cardinality-based inclusion measures for ordinary sets introduced in Chapter 2. In Section 7.3 the $T_{\mathbf{L}}$-transitive members are identified.

7.2 A parametric family of fuzzy inclusion measures

Also inclusion measurement does not restrict to binary vectors only, but is rather applied with vectors whose components are scaled to the unit interval. Therefore, we need fuzzy inclusion measures. We define a fuzzy inclusion measure as follows:

Definition 7.1 *A fuzzy inclusion measure I is a binary fuzzy relation on $\mathcal{F}(X)$ such that*

$$(\forall A, B \in \mathcal{P}(X))(A \subseteq B \Rightarrow I(A,B) = 1)\,.$$

Having introduced a fuzzification rule for the cardinality of a fuzzy set and translated classical set operations in Chapter 3, we are now able to fuzzify the parametric family of inclusion measures (2.5). Therefore, we rewrite the expression (2.5) in terms of intersections only,

$$\begin{aligned} &I(A,B) \\ &= \frac{x(|A| - |A\cap B|) + x'(|B| - |A\cap B|) + y|A\cap B| + z(n - |A| - |B| + |A\cap B|)}{x'(|A| + |B| - 2|A\cap B|) + y|A\cap B| + z(n - |A| - |B| + |A\cap B|)}\,. \end{aligned}$$

Consider two fuzzy sets A and B in a finite universe $X = \{x_1, \ldots, x_n\}$, then we fuzzify the above expression into

$$I(A,B) = \frac{x(a-u) + x'(b-u) + y\,u + z(n-a-b+u)}{x'(a+b-2u) + y\,u + z(n-a-b+u)}\,, \tag{7.1}$$

with $a = \sum_{i=1}^{n} A(x_i)$, $b = \sum_{i=1}^{n} B(x_i)$ and $u = \sum_{i=1}^{n} Q(A(x_i), B(x_i))$, where Q denotes a commutative quasi-copula.

Again, remark that in order to guarantee that $I(A,B) \in [0,1]$, we need to impose the following restriction: $0 \le x \le x'$. Analogously to

Inclusion measure	$T_{\mathbf{M}}$	$T_{\mathbf{P}}$
I_1	$\frac{\sum_{i=1}^{n} B(x_i) - \sum_{i=1}^{n} \min(A(x_i),B(x_i))}{\sum_{i=1}^{n} A(x_i)+B(x_i)-2\min(A(x_i),B(x_i))}$	$\frac{\sum_{i=1}^{n} B(x_i) - \sum_{i=1}^{n} A(x_i)B(x_i)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-2A(x_i)B(x_i)}$
I_2	$\frac{n-\sum_{i=1}^{n} A(x_i)}{n-\sum_{i=1}^{n} \min(A(x_i),B(x_i))}$	$\frac{n-\sum_{i=1}^{n} A(x_i)}{n-\sum_{i=1}^{n} A(x_i)B(x_i)}$
I_3	$\frac{\sum_{i=1}^{n} B(x_i)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-\min(A(x_i),B(x_i))}$	$\frac{\sum_{i=1}^{n} B(x_i)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-A(x_i)B(x_i)}$
I_4	$\frac{n-\sum_{i=1}^{n} A(x_i)+\sum_{i=1}^{n} \min(A(x_i),B(x_i))}{n}$	$\frac{n-\sum_{i=1}^{n} A(x_i)+\sum_{i=1}^{n} A(x_i)B(x_i)}{n}$

Table 7.1: Fuzzy inclusion measures when using $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$.

Inclusion measure	$T_{\mathbf{L}}$
I_1	$\frac{\sum_{i=1}^{n} B(x_i) - \sum_{i=1}^{n} \max(A(x_i)+B(x_i)-1,0)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-2\max(A(x_i)+B(x_i)-1,0)}$
I_2	$\frac{n-\sum_{i=1}^{n} A(x_i)}{n-\sum_{i=1}^{n} \max(A(x_i)+B(x_i)-1,0)}$
I_3	$\frac{\sum_{i=1}^{n} B(x_i)}{\sum_{i=1}^{n} A(x_i)+B(x_i)-\max(A(x_i)+B(x_i)-1,0)}$
I_4	$\frac{n-\sum_{i=1}^{n} A(x_i)+\sum_{i=1}^{n} \max(A(x_i)+B(x_i)-1,0)}{n}$

Table 7.1: Fuzzy inclusion measures when using $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ (continued).

the parametric family of inclusion measures for ordinary sets, the case $x = x'$ leads to trivial measures taking value 1 only, and therefore we consider from here on $0 \leq x < x'$. In Table 7.1, we summarize the fuzzy similarity measures (that are a member of family (7.1) and whose crisp counterparts can be found in Table 2.7) for some commonly used quasi-copulas, $T_{\mathbf{M}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$.

7.3 T-transitive members

First, we identify the $T_{\mathbf{L}}$-transitive members of family (7.1). Next, we fuzzify inclusion measure (2.19) and show that it is $T_{\mathbf{P}}$-transitive. These results are in line with the ones obtained for the family of crisp inclusion measures.

7.3.1 Łukasiewicz-transitive members

Let us recall that a fuzzy inclusion measure is $T_{\mathbf{L}}$-transitive if the following inequality is fulfilled:

$$I(A,C) - I(A,B) - I(B,C) + 1 \geq 0\,. \tag{7.2}$$

Theorem 7.1 *The $T_{\mathbf{L}}$-transitive members of family (7.1) of fuzzy inclusion measures are for any commutative quasi-copula that satisfies I_3^3 characterized by:*

$$x' \geq \max(y, z)\,. \tag{7.3}$$

Proof. In Theorem 2.4 we proved that inequality (7.2) holds for all n and all ordinary sets A, B and C if and only if the parameters satisfy conditions (7.3). For fuzzy sets A, B and C, the left-hand side of inequality (7.2) is a homogeneous function of $|A|$, $|B|$, $|C|$, $|A\cap B|$, $|A\cap C|$, $|B \cap C|$ and $|X|$. When inequality I_3^3 is fulfilled (inequalities I_2^1 and I_3^2 are satisfied for any commutative quasi-copula), we can use Theorem 5.3 to conclude that inequality (7.2) also holds for all fuzzy sets A, B and C under the same parameter conditions (7.3). □

Theorem 7.2 *The $T_{\mathbf{L}}$-transitive members of family (7.1) of fuzzy inclusion measures with $z = 0$ are for any commutative quasi-copula characterized by parameter conditions (7.3).*

Proof. When $z = 0$, the homogeneous function on the left-hand side of (7.2) is independent of $|X|$ and using Theorem 5.4 we can conclude that inequality (7.2) also holds for all fuzzy sets A, B and C under the same parameter conditions (7.3). □

Using the latter theorem, we can conclude that the fuzzified versions of the following inclusion measures

$$I_1(A,B) = \frac{|B \setminus A|}{|A \triangle B|}\,, \quad I_4(A,B) = \frac{|(A \setminus B)^c|}{n}\,,$$

which read as

$$\begin{aligned} I_1(A,B) &= \frac{\sum_{i=1}^n B(x_i) - Q(A(x_i),B(x_i))}{\sum_{i=1}^n A(x_i) + B(x_i) - 2Q(A(x_i),B(x_i))}\,, \\ I_4(A,B) &= 1 - \frac{\sum_{i=1}^n A(x_i) - Q(A(x_i),B(x_i))}{n}\,, \end{aligned}$$

respectively, are $T_{\mathbf{L}}$-transitive, since both are members of family (7.1) with $z = 0$.

However, with Theorem 7.1 we are not able to guarantee the $T_{\mathbf{L}}$-transitivity of other fuzzy inclusion measures when we model the intersection of two fuzzy sets by a commutative quasi-copula which does not satisfy Bell-type inequality I_3^3. To solve this problem, we provide a third theorem. It requires a direct proof and has nothing to do with the meta-theorems discussed in Chapter 5. We first prove the following lemma:

Lemma 7.1 *The following identity is always fulfilled:*

$$\begin{aligned} &(b-u-w+v)[n(b+2c-u)-ac-bw] + n(c+u-v-w)(a-u) \\ &+ (n-b)(a-u)(c-w) + (n-v)(b-u)(b-w) \\ &+ 2(c-v)(b-w)(a-u) + 2n(a-u)(b-w) \\ &+ (b-w)(bw-uv) \geq 0\,. \end{aligned} \tag{7.4}$$

Proof. Since Bell inequalities I_2^1 and I_3^2 are fulfilled for any commutative quasi-copula, we already know that

$$\begin{aligned} (n-b)(a-u)(c-w) &\geq 0\,, \\ (n-v)(b-u)(b-w) &\geq 0\,, \\ 2(c-v)(b-w)(a-u) &\geq 0\,, \\ n(c+u-v-w)(a-u) &\geq 0\,, \end{aligned}$$

and $b+v-u-w \geq 0$. First we prove that $n(b+2c-u)-ac-bw$ is also positive. Since $b-u \geq w-v$, we have that

$$\begin{aligned}
& n(b+2c-u)-ac-bw \\
\geq\ & n(2c+w-v)-ac-bw \\
=\ & c(n-a)+n(c-v)+w(n-b) \\
\geq\ & 0\,.
\end{aligned}$$

Next, we prove that $2n(a-u)(b-w)+(b-w)(bw-uv) \geq 0$ is always fulfilled:

$$\begin{aligned}
& 2n(a-u)(b-w)+(b-w)(bw-uv) \\
=\ & n(a-u)(b-w)+(b-w)(n(a-u)+bw-uv) \\
\geq\ & n(a-u)(b-w)+(b-w)(n(a-u)+uw-uv) \\
\geq\ & n(a-u)(b-w)+(b-w)(n(a-u)-u(a-u)) \\
=\ & n(a-u)(b-w)+(b-w)(n-u)(a-u) \\
\geq\ & 0\,.
\end{aligned}$$

Therefore inequality (7.4) is always fulfilled. □

Theorem 7.3 *The $T_{\mathbf{L}}$-transitive members of family (7.1) of fuzzy inclusion measures are for any commutative quasi-copula characterized by parameter conditions (7.3).*

Proof. In order to identify the conditions on the parameters x, x', y and z in (7.1), we have to verify when the following inequality is fulfilled:

$$\begin{aligned}
(x'-x)\Bigg(&-\frac{a-v}{x'(a+c-2v)+y\,v+z(n-a-c+v)} \\
&+\frac{a-u}{x'(a+b-2u)+y\,u+z(n-a-b+u)} \\
&+\frac{b-w}{x'(b+c-2w)+y\,w+z(n-b-c+w)}\Bigg) \geq 0\,.
\end{aligned} \tag{7.5}$$

Since $x' > x$ we can omit the factor $x'-x$.

Case $z=0$.

Substituting $z=0$ in inequality (7.5), converting the fractions such that they have a common denominator and omitting this denominator since it is always positive, the following inequality should be verified:

$$K(x',y) = \alpha y^2 + \beta y + \gamma \geq 0\,,$$

with

$$\begin{aligned}
\alpha &= -auw + avw - uvw + buv\,,\\
\beta &= x'(2abv - abw - acu + acv + acw + 2auw - 4avw + b^2v\\
&\quad + bcu - 4buv - 2cuw + 4uvw)\,,\\
\gamma &= x'^2(a^2b + ac^2 - a^2w + abc - abu - 3abv + abw + acu - acv\\
&\quad - 3acw + 4avw + b^2c - b^2v - 3bcu + bcv - bcw + 4buv\\
&\quad - c^2u + 4cuw - 4uvw)\,.
\end{aligned}$$

It is easy to prove, using Theorem 5.4, that $K(x',0) = \gamma \geq 0$ and $K(x',x') = x'^2(\alpha+\beta+\gamma) \geq 0$. To apply Lemma 6.2, we still have to verify that $2\gamma + \beta x' \geq 0$. Again, using Theorem 5.4 it is easy to see that this inequality is fulfilled. Remark that since every commutative quasi-copula satisfies Bell inequality I_3^2, applying Theorem 5.4 does not impose any limitations to our theorem.

We can conclude that the condition

$$x' \geq y \,\wedge\, z = 0$$

is a sufficient condition for inequality (7.5) to hold.

Case $y = 0$**.**
Substituting $z = 0$ in inequality (7.5), we obtain a quadratic function $K(x',z)$ in z. In the same way as the previous case, we can prove that $K(x',z) \geq 0$ for any $z \in [0,x']$. Therefore, the condition

$$x' \geq z \,\wedge\, y = 0$$

is a sufficient condition for inequality (7.5) to hold.

Case $y \neq 0,\ z \neq 0$**.**
In this case we have to verify that the following inequality

$$K(x',y,z) = c_1y^2 + c_2yz + c_3z^2 + c_4y + c_5z + c_6 \geq 0\,, \qquad (7.6)$$

with c_4 and c_5 proportional to x' and c_6 proportional to x'^2, is fulfilled for any $y, z \in [0,x']$.

Consider the left part of inequality (7.6) as a quadratic function in y. Since $K(x',0,z) \geq 0$ (this was proven in the case $y = 0$) and $K(x',x',z) \geq 0$ (again, we use Lemma 6.2: since $K(x',x',0) \geq 0$, which was proven in the case $z = 0$, and $K(x',x',x') = n^2(b-u-w+v) \geq 0$, it follows that $K(x', x', z) \geq 0$ for any $z \in [0, x']$), we can use Lemma 6.2 to prove that inequality (7.6) is fulfilled for any $y \in [0, x']$. Therefore,

$$2(c_3 z^2 + c_5 z + c_6) + (c_2 z + c_4)x' \geq 0\,,$$

or equivalently,

$$2c_3 z^2 + (2c_5 + c_2 x') + 2c_6 + c_4 x' \geq 0$$

should be fulfilled for any $z \in [0, x']$. Again, we invoke Lemma 6.2. Since $K(x', y, 0) \geq 0$ and $K(x', y, x') \geq 0$, we should only verify that the following inequality

$$c_2 x'^2 + 2(c_4 + c_5)x' + 4c_6 \geq 0 \tag{7.7}$$

is fulfilled for any x'. Since $x' \geq 0$, inequality (7.7) is equivalent to

$$\begin{aligned}
&(b-u-w+v)[n(b+2c-u) - ac - bw] + n(c+u-v-w)(a-u)\\
&+ (n-b)(a-u)(c-w) + (n-v)(b-u)(b-w)\\
&+ 2(c-v)(b-w)(a-u) + 2n(a-u)(b-w)\\
&+ (b-w)(bw-uv) \geq 0\,.
\end{aligned}$$

It was already proven in Lemma 7.1 that the latter inequality is always satisfied, therefore, inequality (7.7) is fulfilled for any x'. Consequently, also inequality (7.6) is satisfied for any $y, z \in [0, x']$. We can conclude that

$$x' > x \,\wedge\, x' \geq \max(y, z)\,,$$

are sufficient conditions such that the members of family (7.1) of fuzzy inclusion measure are $T_{\mathbf{L}}$-transitive.

The proof that these conditions are also necessary conditions is completely analogously to that of Theorem 2.4. □

Corollary 7.1 *The $T_{\mathbf{L}}$-transitive members of family (7.1) of fuzzy inclusion measures are for any Frank t-norm characterized by parameter conditions (7.3).*

This corollary immediately follows from Theorem 7.3. It only follows from Theorem 7.1 for $\lambda \leq 9 + 4\sqrt{5}$ or from Theorem 7.2 when $z = 0$.

7.3.2 Product-transitive members

Since it is not possible to establish a set of conditions that are at the same time necessary and sufficient such that the product-transitive members of family (2.5) are characterized, we are not able to characterize the $T_{\mathbf{P}}$-transitive members of family (7.1). However, we can fuzzify inclusion measure (2.19), which reads as

$$I(A,B) = \frac{5(a-u)+10(b-u)+u}{10(a+b-2u)+u}\,, \tag{7.8}$$

and prove the following proposition.

Proposition 7.1 *Fuzzy inclusion measure (7.8) is $T_{\mathbf{P}}$-transitive.*

Proof. We have to verify when the following inequality is fulfilled:

$$\begin{aligned}&\left(\frac{5(a-u)+10(b-u)+u}{10(a+b-2u)+u} \cdot \frac{5(b-w)+10(c-w)+w}{10(b+c-2w)+w}\right)\\ \leq\ & \frac{5(a-v)+10(c-v)+v}{10(a+c-2v)+v}\,.\end{aligned} \tag{7.9}$$

Since inequality (7.9) is independent of n, we can use Theorem 5.4 to conclude that inequality (7.9) is fulfilled for any commutative quasi-copula and hence inclusion measure (7.8) is $T_{\mathbf{P}}$-transitive. □

7.4 Conclusions

In this chapter, we have fuzzified the parametric family of inclusion measures for ordinary sets, which was introduced in Chapter 2. To that end, we have modelled the intersection of two fuzzy sets by a commutative quasi-copula. The main result of this chapter is that $T_{\mathbf{L}}$-transitivity, is preserved along this fuzzification process. Unfortunately, we were not able to establish a set of conditions that are at the same time necessary and sufficient such that the product-transitive members of family (2.5) are characterized. Therefore, we were not able to characterize the $T_{\mathbf{P}}$-transitive members of family (7.1) either.

Chapter 8

Applications

8.1 Similarity measures in clustering tasks

8.1.1 Introduction

Many machine learning tasks require similarity measures that estimate likeness between observations. Similarity measures are particularly important for clustering algorithms that depend on estimates of the distance between data points. However, standard measures such as the Euclidean distance or the Jaccard coefficient often fail to capture an appropriate notion of similarity for a particular domain or dataset. This problem can be alleviated by employing a family of fuzzy similarity measures such that a whole range of similarity measures becomes available.

Clustering can be roughly defined as the problem of partitioning a data set into disjoint groups so that observations belonging to the same cluster are similar, while observations belonging to different clusters are dissimilar. Clustering has been widely studied for several decades and a variety of algorithms exists. Similarity measures, however, are central to the clustering problem regardless of the particular algorithm used, since all of them utilize similarity measures between observations and clusters or between individual observations [59].

8.1.2 A small clustering example

In this section, we want to illustrate, with a small example, that using different similarity measures when clustering a data set can have a strong impact. For this experiment, we have used the wine data set, on-line available from the UCI Machine Learning Repository[1]. This dataset contains information about 178 bottles of wine originating from three different winegrowers. A desirable clustering consist in three clusters, each corresponding to a particular winegrower.

Many algorithms have been proposed for cluster analysis. Hierarchical clustering is a similarity-based bottom-up clustering technique in which at the beginning every data point forms a cluster of its own. Then, the algorithm iterates over the step that merges the two most similar clusters still available, until one arrives at a universal cluster that contains all data points. In our experiments, we use one particular strategy to calculate the similarity between clusters: complete linkage. Note that also single linkage and average linkage clustering do exist. These three strategies measure the dissimilarity between two non-trivial clusters in different ways. If n_r is the number of objects in cluster r and n_s is the number of objects in cluster s, and x_{ri} is the ith object in cluster r, for any $i \in \{1, \ldots, n_r\}$, and x_{sj} is the jth object in cluster s, for any $j \in \{1, \ldots, n_s\}$, then complete linkage defines the distance between the clusters r and s as the maximum distance between them, i.e.

$$d(r, s) = \max_{i,j}(dist(x_{ri}, x_{sj}))\,,$$

while single linkage uses the smallest distance between objects in the two clusters and average linkage uses the average distance between all pairs of objects in cluster r and cluster s.

The wine data set consists of 178 instances, each with 14 fields. The first field indicates to which class (or equivalently, to which winegrower) an instance belongs, while the next 13 fields are the feature values[2], which are real numbers. First, the 13 feature values were transformed into values belonging to the unit interval such that the family of similarity measures (6.1) can be employed. Therefore, we determined for each feature j (with $j = 1, \ldots, 13$) its maximum value u_j and its minimum value l_j. Then, the feature values x_{ij}, for any $i = 1, \ldots, 178$ and for any $j = 1, \ldots, 13$, are transformed into

$$z_{ij} = \frac{x_{ij} - l_j}{u_j - l_j}$$

such that $z_{ij} \in [0, 1]$.

[1] http://www.ics.uci.edu/~mlearn/MLRepository.html

[2] The attribute values are alcohol, malic acid, ash, alcalinity of ash, magnesium, phenols, flavonoids, nonflavonoid phenols, proanthocyanins, color intensity, hue, dilution and proline.

Dissimilarity coefficient	$D(A,B) =$
Jaccard	$1 - \frac{\sum_{i=1}^n \min(A(x_i),B(x_i))}{\sum_{i=1}^n \max(A(x_i),B(x_i))}$
Simple matching	$\frac{\sum_{i=1}^n A(x_i)+\sum_{i=1}^n b_i-2\sum_{i=1}^n \min(A(x_i),B(x_i))}{n}$
Rogers and Tanimoto	$2\frac{\sum_{i=1}^n A(x_i)+\sum_{i=1}^n B(x_i)-2\sum_{i=1}^n \min(A(x_i),B(x_i))}{n+\sum_{i=1}^n A(x_i)+\sum_{i=1}^n B(x_i)-2\sum_{i=1}^n \min(A(x_i),B(x_i))}$
Sokal and Sneath	$\frac{\sum_{i=1}^n A(x_i)+\sum_{i=1}^n A(x_i)-2\sum_{i=1}^n \min(A(x_i),B(x_i))}{2n-\sum_{i=1}^n A(x_i)-\sum_{i=1}^n B(x_i)+2\sum_{i=1}^n \min(A(x_i),B(x_i))}$
Euclidean distance	$\left(\sum_{i=1}^n (A(x_i)-B(x_i))^2\right)^{1/2}$

Table 8.1: Overview of dissimilarity measures.

Dissimilarity coefficient	Number of misclassified instances
Jaccard	75
Simple matching	10
Rogers and Tanimoto	10
Sokal and Sneath	10
Euclidean distance	12
Euclidean distance (original data set)	58

Table 8.2: Overview of misclassified instances using different dissimilarity measures.

Next, we constructed the 178×178 matrix of pairwise distances using the following dissimilarity measures: Euclidean distance, the Jaccard dissimilarity coefficient, the simple matching dissimilarity coefficient, the Rogers and Tanimoto dissimilarity coefficient and the Sokal and Sneath dissimilarity coefficient, which are clarified in Table 8.1. We also constructed the 178×178 matrix of pairwise distances using the Euclidean distance on the original data set (i.e. the last 13 fields of the 178 instances without transforming them to the unit interval).

We performed a complete linkage clustering (using three clusters) on these distance matrices and verified how many instances were misclassified. We conclude that using the simple matching dissimilarity coefficient, the Rogers and Tanimoto dissimilarity coefficient or the Sokal and Sneath dissimilarity coefficient the number of misclassified instances (which is equal to 10) is much lower than using the standard dissimilarity measures like Euclidean distance (on the original data set) and the Jaccard dissimilarity measure. In the latter case, the number of misclassified instances is equal to 58 and 78, respectively. We also conclude that using the Euclidean distance on the transformed data set the number of misclassified instances (which is equal to 12) is much lower than using the Euclidean distance on the original data set (which is equal to 12). The results are summarized in Table 8.2. All experiments were done in MATLAB using the Statistics Toolbox.

This small experiment shows that the family of fuzzy similarity measures (6.1) can be of great value when choosing a dissimilarity measure in a clustering task. Although we did not perform other experiments, we are convinced that this family of fuzzy similarity measures will be useful in any other real-world application where (dis)similarity is involved.

8.2 Inclusion measures for leaf-labelled trees

8.2.1 Introduction

Whether biologists are interested in the history of life (the Tree of Life), or in using evolutionary relationships to analyze biological data from other fields (e.g. analyzing rapidly mutating viruses such as HIV, identifying species that may disrupt ecosystems, etc.), phylogenetic trees provide a comparative framework for understanding and interpreting these biological data. Many methods to reconstruct these trees have already been developed; the three main schools of reconstruction methods are maximum likelihood methods, maximum parsimony methods and distance-based methods.

A phylogenetic tree is usually represented by a leaf-labelled tree, where the internal nodes refer to hypothetical ancestors, the leaves are labelled by the taxonomic unit (genes, species, populations, individuals) and the branches define the relationship between the taxonomic units in terms of descent and ancestry. The tree can be rooted or unrooted depending on the availability of sufficient information to decide the orientation of the evolution. If the tree is rooted, then the root is a common ancestor for all the species in the tree. A phylogenetic tree is unordered, i.e. for rooted trees one does not distinguish between different orderings of the children of a node and for unrooted trees one does not distinguish between different orderings between the neighbours. Finally, they can be binary as well as non-binary.

The comparison of phylogenetic trees is a fundamental problem in biology. Different evolutionary hypotheses arise when different reconstruction methods are applied to the same set of data, or when a single method is applied to different data sets. Several similarity metrics between phylogenetic trees are currently in use. First, we will present an overview of major approaches to tree comparison.

According to Boorman and Olivier [9], there are two basic approaches to constructing tree metrics. The first one is to formalize the concept of 'transformation of a tree' and to define a metric on trees as the least number of moves necessary to transform one tree into another. An example of such a metric is the nearest neighbour interchange (nni) metric [85]. The similarity of two binary trees is represented by counting the minimum number of nni operations required to change one tree into another. An nni operation swaps two subtrees that are separated by an internal edge. Computing the nni distance is NP-complete [18]. A more general metric than the nni metric, the partition metric, is due to Penny and Hendy [65]. Trees are compared by counting the number of edges in one tree for which there is no equivalent edge on the other tree (and vice versa). The partition metric is suitable for both rooted and unrooted trees, as well as binary and non-binary trees. Another example which also fits in this particular class is the tree editing distance. The distance between two trees is then defined as the cost of a sequence of edit operations (delete or insert a node or modify a label of the node) to transform one tree into the other one [73].

The second approach consists of representing a tree in terms of simpler structures for which adequate metrics are available. Estabrook et al. [31] propose various kinds of similarity and dissimilarity measures to compare unrooted phylogenetic trees for the same collection of species. Their ideas are based on the fact that four species (a quartet) is the smallest number about which distinct statements concerning branching patterns in unrooted trees can be made. Bryant et al. [12] describe an $O(n^2)$ algorithm that computes the quartet distance between two phylogenetic trees.

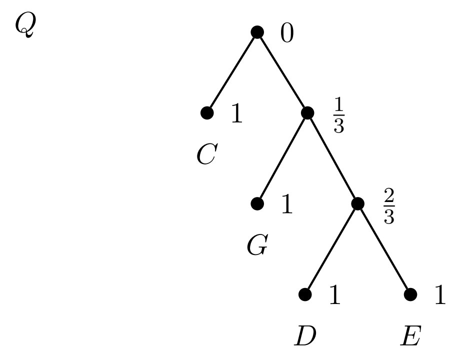
A third approach can be added to the previous ones. In this approach, the similarity between phylogenetic trees is based on subtree similarity. Zhong et al. [100] have developed a general comparison methodology between different leaf-labelled trees. They use a similarity measure for ordinary sets to compare first pairs of subtrees (which are simply reduced to their leaf node sets) and they further propose a corresponding algorithm, the so-called webbing matrix method, for measuring overall similarity between leaf-labelled trees.

With the growing number of phylogenetic trees available, the need of managing phylogenetic databases is great. The comparison of such trees is one thing, but on the other hand also methods for searching and retrieval in these databases are needed. In [83], an inclusion measure for phylogenetic trees (TreeRank) was proposed and successfully implemented in the phylogenetic information system TreeBASE. This method can only be applied to rooted, unordered trees and is based on the additive distance matrix of a phylogenetic tree and a *data tree reduction* technique where nodes that are not in common to the trees to be compared are removed. Other approaches in the literature are mainly based on the minimal number of operations needed to transform one tree into another: a tree T_1 is included in T_2 if and only if T_1 can be obtained only by deleting nodes from T_2 (e.g. [50]).

8.2.2 Tree inclusion

In this section, we propose a method based on fuzzy inclusion measures for attributing a degree of inclusion of one leaf-labelled tree in another one. Our method aims at representing a tree in terms of simpler structures. Therefore, we attribute weights to the nodes of a tree and associate a symmetric matrix with the tree. We further assume that the tree is an unordered rooted tree with internal nodes possessing at least two children.

First, we attribute weights to the nodes of a tree as follows. Among all paths in the tree from the root to a leaf node and passing through a given node, we select a path with maximum length. If that path has length q (i.e. it contains q edges) and the given node lies p edges away from the root, the weight p/q is attributed to that node. Note that the root has weight 0 and all the (labelled) leaf nodes have weight 1. This procedure of attributing weights to tree nodes is illustrated in Figures 8.1 and 8.2 on two example trees: a query tree Q and a data tree D. The common leaf node label set is $\mathcal{L} = \{C, D, E\}$, the labels A, B, F are missing in Q and the label G is missing in D. If we order the leaf node labels in a standard but otherwise arbitrary way (here we choose the alphabetical order), we can unambiguously associate with any tree T a symmetric matrix $\hat{T}$ indexed by the ordered leaf labels. For any $X, Y \in \mathcal{L}$, the element $\hat{T}_{X,Y}$ equals the weight of the least common ancestor in the tree T of the two leaf nodes with respective labels X and Y. With any label that is not present in the tree, we associate a row and column of all zeros in the matrix. Finally, we remark that the tree is nothing else but the partition tree associated with the matrix (ignoring the zero-rows and zero-columns), and therefore the matrix itself is $T_{\mathbf{M}}$-transitive. In Figures 8.1 and 8.2 the matrices corresponding to Q and D are shown under the trees.



$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2/3 & 0 & 1/3 \\ 0 & 0 & 0 & 2/3 & 1 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1 \end{pmatrix}$$

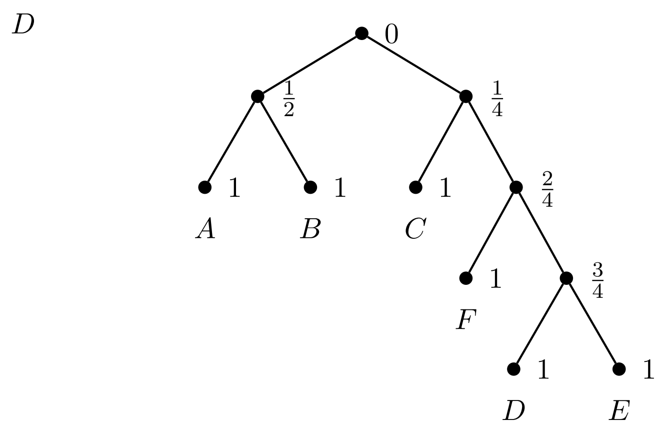
Figure 8.1: A query tree Q with its corresponding matrix $\hat{Q}$.

An overall coefficient of inclusion between two trees Q and D is obtained by first selecting a particular fuzzy inclusion measure I and then calculating $I(\hat{Q}, \hat{D})$, where the matrices are interpreted as fuzzy sets in the universe $\mathcal{L}^2$. For the previous example, we obtain with the members of the family of fuzzy inclusion measures (7.1) whereby the intersection of two fuzzy sets is modelled by $T_{\mathbf{M}}$ (note that these inclusion measures are also listed in Table 7.1):

$$\begin{aligned} I_1(\hat{Q}, \hat{D}) &= 0.7544\,, \\ I_2(\hat{Q}, \hat{D}) &= 0.8313\,, \\ I_3(\hat{Q}, \hat{D}) &= 0.9478\,, \\ I_4(\hat{Q}, \hat{D}) &= 0.9524\,. \end{aligned}$$

Remark that one could opt to calculate $I(\hat{Q}, \hat{D})$ using upper triangle matrices only, but this does not result in any major differences.

One of the methods of measuring inclusion available in the literature is called TreeRank [83]. Consider a query and a data tree, then this method also builds a simpler structure from those trees, the so-called UpDown matrix, based on the number of up and down operations be-



$$
\begin{pmatrix}
1 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
1/2 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1/4 & 1/4 & 1/4 & 0 \\
0 & 0 & 1/4 & 1 & 3/4 & 1/4 & 0 \\
0 & 0 & 1/4 & 3/4 & 1 & 1/2 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

Figure 8.2: A data tree D with its corresponding matrix $\hat{D}$.

tween two nodes in a tree.

Example 8.1 Consider the query tree Q form Figure 8.1 and the data tree D from Figure 8.2. Then the UpDownmatrices U_Q and U_D are given by

$$U_Q = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 3 & 2 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix} \text{ and } U_D = \begin{pmatrix} 0 & 1 & 2 & 2 & 2 & 2 \\ 1 & 0 & 2 & 2 & 2 & 2 \\ 2 & 2 & 0 & 1 & 1 & 1 \\ 4 & 4 & 3 & 0 & 1 & 2 \\ 4 & 4 & 3 & 1 & 0 & 2 \\ 3 & 3 & 2 & 1 & 1 & 0 \end{pmatrix},$$

respectively.

A data reduction tree technique is incorporated in their method: the data tree results in a reduced data tree by removing nodes in such a way that the reduced data tree only possesses nodes that are common to both the query and the data tree.

Example 8.2 Consider again the query tree Q and the data tree D from Figures 8.1 and 8.2, respectively. The common leaf label node set is $\mathcal{L} = \{C, D, E\}$. Then the data tree D is reduced to the tree D' as depicted in Figure 8.3. Note that also the UpDownmatrix U_D changes into $U_{D'}$, with

$$U_{D'} = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

Then the TreeRank score form Q to D' is calculated as follows. Let V_Q be the set of labeled nodes in Q and let $V_{D'}$ be the set of labeled nodes in D'. Let $\mathcal{L}$ denote the common leaf label node set, i.e. $\mathcal{L} = V_Q \cap V_{D'}$ and let $\mathcal{J}$ denote the leaf labels which are present in Q, but not in D', i.e. $\mathcal{J} = V_Q \setminus V_{D'}$. Then the distance from Q to D', denoted as $\text{dist}(Q, D')$, is defined as

$$\text{dist}(Q, D') = \sum_{u \in \mathcal{L}} \sum_{v \in \mathcal{L}} |U_Q(u, v) - U_{D'}(u, v)| + \sum_{u \in \mathcal{J}} \sum_{v \in \mathcal{J}} U_Q(u, v).$$

The TreeRank score from Q to D, is calculated by

$$\text{TreeRank}(Q, D) = 1 - \frac{\text{dist}(Q, D')}{\sum_{u \in V_Q} \sum_{v \in V_Q} U_Q(u, v)}.$$

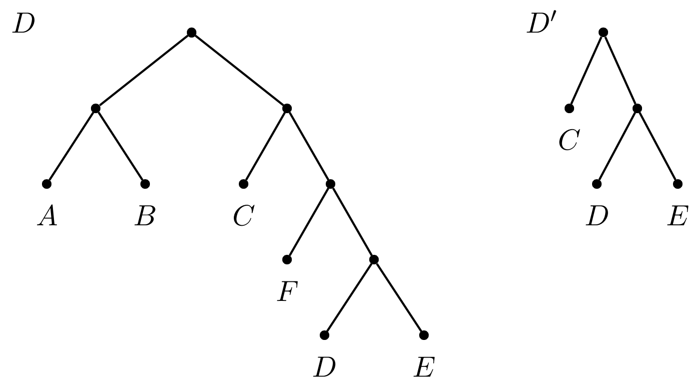


Figure 8.3: Example showing how tree D is reduced to D' using the data tree reduction technique.

The advantage of our method is that no changes (neither delete, nor insert operations) have to be performed on any of the trees involved in order to obtain a degree of inclusion. Another difference between our method and TreeRank is that our method takes more into account the label context when the topological relationship of a query tree is found to be similar to that in the data tree. Take the previous example, then TreeRank gives a score 1, which means that, according to TreeRank, tree Q is considered to be *fully included* in tree D. Nonetheless, it is easy to see that the topological relationship of both trees is very similar, but the leafs labelled G and F differ from each other, which is not taken into account by TreeRank.

Another approach in the literature is the use of delete, insert and relabelling operations to transform one tree into another. The 'tree inclusion problem' is then defined as follows: T_1 is included in T_2 if and only if T_1 can be obtained by deleting nodes from T_2 [50]. Obviously, this 'is included in' relation is transitive, yet crisp. Our method of measuring inclusion leaves room for nuance as it is $[0,1]$-valued; moreover, $T_{\mathbf{L}}$- or $T_{\mathbf{P}}$-transitivity can be guaranteed as well.

In our approach $I(\hat{Q},\hat{D}) = 1$ if and only if $\hat{Q}_{XY} \leq \hat{D}_{XY}$ for all $(X,Y) \in \mathcal{L}_Q^2$, which implies in particular that $\mathcal{L}_Q \subseteq \mathcal{L}_D$. It is important to realize that $I(\hat{Q},\hat{D}) = 1$ does not necessarily mean that tree Q 'is included in' tree D. Also, if tree Q 'is included in' tree D, then we do not necessarily have that $I(\hat{Q},\hat{D}) = 1$. The following example will make things clear. Consider the trees as in Figure 8.4. Then $I_1(\hat{Q}_1,\hat{D}_1) = 0.9856$, while tree Q_1 'is included in' tree D_1. On the

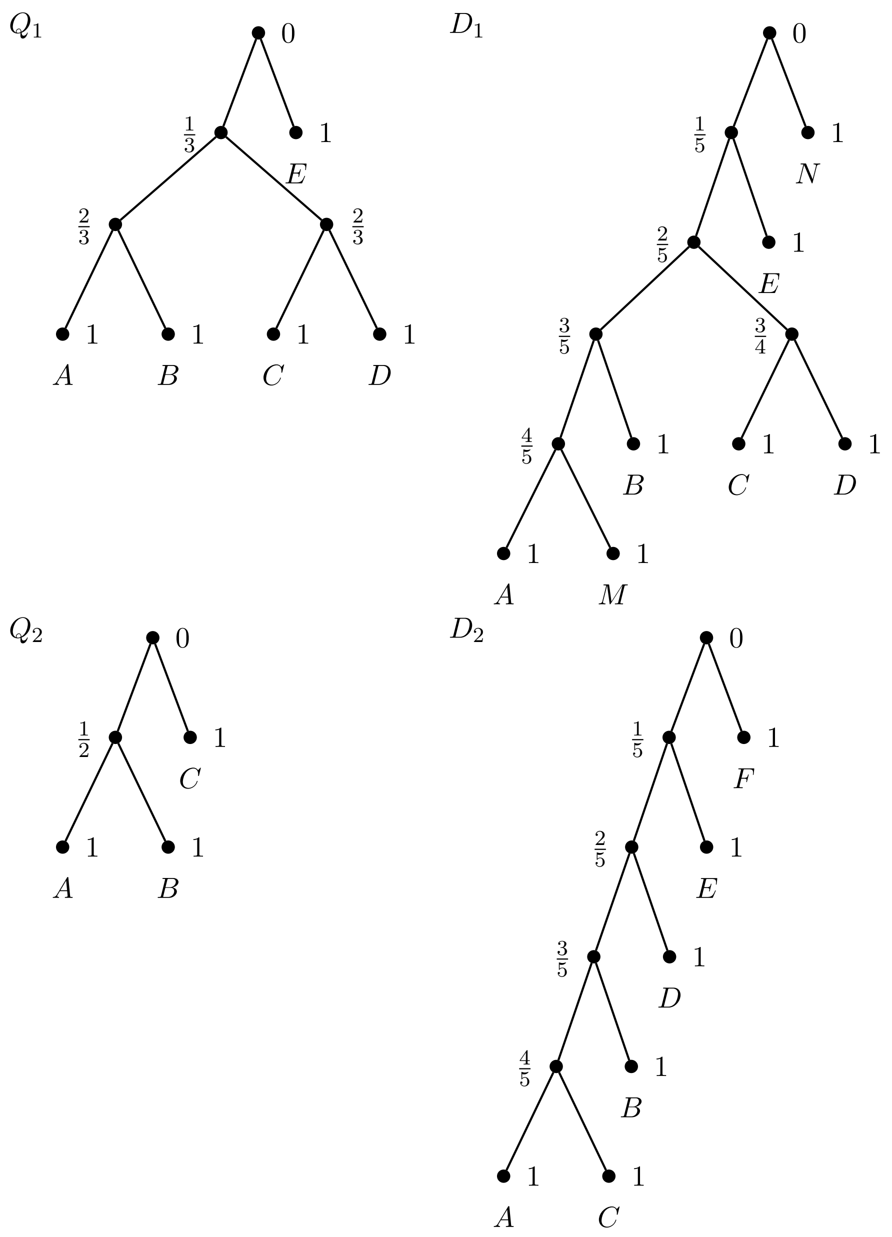


Figure 8.4: Two query trees Q_1 and Q_2 and two data trees D_1 and D_2.

other hand, tree Q_2 is certainly 'not included in' tree D_2, while our method yields $I_1(\hat{Q}_2, \hat{D}_2) = 1$. In the latter case, our method is obviously blurred by the larger context of tree D_2.

8.3 Conclusions

In this chapter, we showed that the family of fuzzy similarity measures (6.1) as well as the family of fuzzy inclusion measures (7.1) can be of great value in real-world applications.

Chapter 9

Nederlandstalige samenvatting

In ons dagelijks leven vergelijken we continu dingen met elkaar. We vergelijken personen of objecten en maken voortdurend keuzes, meestal zonder dat we er ons van bewust zijn. Vergelijkbaarheid speelt dus een belangrijke rol. Maar niet alleen in ons alledaagse leven, ook in vele wetenschappelijke domeinen neemt "het vergelijken" een prominente rol in. Het vergelijken van objecten is immers één van de deelaspecten in het proces om informatie uit data te extraheren.

We zullen in dit werk methodes invoeren om te bepalen in welke mate twee objecten gelijk zijn. Deze methodes zullen we aanduiden met de term similariteitsmaten. Maten om objecten met elkaar te vergelijken zijn niet nieuw. In de literatuur vinden we reeds heel wat similariteitsmaten terug die doorheen de jaren geïntroduceerd werden in verschillende domeinen en voor uiteenlopende toepassingen. De meeste similariteitsmaten vinden hun oorsprong in het domein van de numerieke taxonomie. Biologen hadden immers een instrument nodig om de objecten die ze bestudeerden (in dit geval meestal planten of dieren) te vergelijken met elkaar en zo een taxonomie op te stellen. Deze similariteitsmaten berusten op een telprincipe, waarbij een zekere weging gemaakt wordt van gemeenschappelijke en/of ontbrekende kenmerken van de te vergelijken objecten.

Om de reeds bestaande similariteitsmaten te kunnen gebruiken, moeten we de objecten die we willen vergelijken met elkaar, vertalen naar $\{0, 1\}$-vectoren (ook wel binaire vectoren genoemd). Voor deze objecten wordt dan een verzameling van kenmerken x_i, voor $i = 1, \ldots, n$ gedefinieerd (d.i. het universum X) zodat de corresponderende binaire vectoren een weerslag zijn van de aan- of afwezigheid van elk van die kenmerken. Een kenmerk dat afwezig is, wordt dan voorgesteld door een 0 in een binaire vector, terwijl een 1 staat voor de aanwezigheid van dat kenmerk. Zo'n $\{0,1\}$-vector kan echter geïdentificeerd worden met een scherpe verzameling op de volgende manier: kenmerk x_i behoort tot deze verzameling als en slechts als op plaats i in de vector een 1 staat. Als gevolg hiervan kunnen de reeds bestaande maten eenvoudig omgezet worden naar maten voor scherpe verzamelingen, louter gebaseerd op de cardinaliteiten van de doorsnede, de unie of het verschil van deze (scherpe) verzamelingen. Similariteitsmaten worden vervolgens gedefinieerd als reflexieve, symmetrische (binaire) vaagrelaties over de machtsverzameling van het universum X. Enkele bestaande families, zoals Tversky's contrast model en de families van Gower en Legendre worden op deze manier vertaald naar families gebaseerd op cardinaliteiten van scherpe verzamelingen.

In dit werk voeren we ook een nieuwe, geparametriseerde familie similariteitsmaten in, die louter gebaseerd is op de cardinaliteiten van scherpe verzamelingen. De meest bekende similariteitsmaten (zoals de Jaccard coëfficiënt, de Dice coëfficiënt, de Simple Matching coëfficiënt en andere) behoren tot deze familie, alsook de twee families geïntroduceerd door Gower en Legendre. Deze familie similariteitsmaten voorziet de lezer van een breed spectrum aan maten en laat hem ook toe nieuwe similariteitsmaten te gebruiken.

Een similariteitsmaat kan gemakkelijk in een dissimilariteitsmaat omgezet worden. Deze laatstgenoemde maten meten het verschil tussen bepaalde objecten en zijn nauw verwant met een afstandsmaat (zoals bijvoorbeeld de Euclidische afstand). Daarom is het sterk aan te raden dat een dissimilariteitsmaat voldoet aan de eigenschappen van een metriek (scheidingseigenschap, symmetrie en de driehoeksongelijkheid). Wanneer uitsluitend identieke koppels onderlinge afstand 0 hebben, dan wordt de dissimilariteitsmaat een pseudometriek genoemd, terwijl wanneer de driehoeksongelijkheid versterkt wordt naar $\max(d(x,y), d(y,z)) \geq d(x,z)$, we te maken hebben met een ultrametriek.

T-transitiviteit (waarbij T een willekeurige driehoeksnorm voorstelt) wordt gezien als één van de belangrijkste eigenschappen die een vaagrelatie kan bezitten. Er bestaat echter een verband tussen T-transitiviteit en een ultrametriek en een pseudometriek: wanneer een similariteitsmaat S min-transitief (resp. Łukasiewicz-transitief) is, is de corresponderende dissimilariteitsmaat $D = 1 - S$ een ultrametriek (resp. pseudometriek) en wanneer een similariteitsmaat S produkt-transitief is, is de corresponderende dissimilariteitsmaat $D = -\log S$ een pseudometriek.

Deze verbanden duiden meteen het belang van T-transitiviteit bij een similariteitsmaat aan (want dan blijft immers de duale metrische interpretatie ook bewaard). Eén van de belangrijkste bijdragen van dit werk, naast het introduceren van een nieuwe parametrische familie similariteitsmaten, is dan ook het identificeren van de parameters zodat we de Łukasiewicz- en product-transitieve leden van deze familie kunnen karakteriseren. Helaas bevat deze familie geen enkel min-transitief lid.

We hebben het hier echter steeds over het vergelijken van objecten. De graad van overeenkomst tussen twee objecten uitdrukken is dan ook het eerste wat in ons opkomt wanneer we het hebben over het vergelijken van objecten. Nochtans is het vergelijken slechts één van de mogelijkheden om naar de verwantschap tussen objecten te kijken. We kunnen ons evengoed afvragen in welk mate het ene object vervat zit in het andere. Het ideale gereedschap om deze probleemstelling aan te pakken is een inclusiemaat.

Alhoewel in de literatuur heel veel artikels in verband met similariteitsmaten kunnen gevonden worden, is het aanbod van publicaties in verband met (scherpe) inclusiematen zeer beperkt. We introduceren dan ook een parametrische familie inclusiematen, gebaseerd op cardinaliteiten, en onderzoeken eveneens voor welke waarden van de parameters de leden van deze familie Łukasiewicz- of product-transitief zijn.

Om deze op cardinaliteiten gebaseerde families van similariteits- of inclusiematen te kunnen gebruiken, moeten we echter nog steeds de objecten die we met elkaar willen vergelijken, reduceren naar $\{0,1\}$-vectoren. Wanneer we echter denken aan louter zwart/wit argumenten om keuzes te maken, dan is het logisch dat deze vereenvoudiging van objecten naar $\{0,1\}$-vectoren ervoor zorgt dat heel wat informatie verloren gaat. Het is meer aannemelijk dat objecten getransformeerd worden naar vectoren met reële coefficiënten, waarbij een hoge waarde van deze coëfficiënt duidt op een kenmerk dat in een hoge mate aanwezig is, terwijl een lage waarde aangeeft dat een kenmerk in mindere mate aanwezig is (zoals bijv. microarray data, gewichten van kernwoorden in documenten, enz.). Elke vector met dergelijke reële coefficiënten kan evenwel herschaald worden naar een vector met coefficiënten in

$[0, 1]$. En zo belanden we bijna automatisch bij de vaagverzamelingenleer.

Een vaagverzameling wordt immers gekarakteriseerd door een $X \to [0, 1]$ afbeelding waarbij $A(x)$ de lidmaatschapsgraad voorstelt van x in A. Wanneer $A(x)$ gelijk is aan 0, behoort x niet tot A, de waarde 1 daarentegen geeft aan dat x volledig tot A behoort (dit gedrag correspondeert dus met scherpe verzamelingen), terwijl een waarde die tussen 0 en 1 ligt een intermediaire graad van lidmaatschap aangeeft. Dus, zoals $\{0, 1\}$-vectoren geïdentificeerd kunnen worden met scherpe verzamelingen, zo kunnen $[0, 1]$-vectoren geïdentificeerd worden met vaagverzamelingen.

We willen in dit werk dan ook een familie vage similariteitsmaten en een familie vage inclusiematen construeren, vertrekkende van de respectievelijke families voor scherpe verzamelingen, zodanig dat de transitiviteitseigenschappen die gelden voor scherpe verzamelingen bewaard blijven. Het is frappant dat we, bij het nagaan van die transitiviteitseigenschappen, steeds dezelfde ongelijkheden aantreffen die moeten vervuld zijn. Deze ongelijkheden lijken formeel op de Bell ongelijkheden.

De Bell ongelijkheden komen voort uit een discussie tussen Einstein en Bohr in het begin van de twintigste eeuw naar aanleiding van de EPR-paradox, die beweerde dat er ofwel acties van op een afstand konden genomen worden ofwel dat de kwantummechanica een incomplete theorie was. Bell formuleerde als antwoord hierop bepaalde ongelijkheden voor correlaties (dit zijn kansen dat twee gebeurtenissen beide plaatsvinden), zodanig dat, als aan deze ongelijkheden was voldaan, er een klassiek waarschijnlijkheidsmodel bestond voor de probabiliteiten die met de uitkomsten van de experimenten gerelateerd zijn.

Aangezien de Bell ongelijkheden gelden voor (klassieke) probabiliteiten, kunnen ze herschreven worden als ongelijkheden voor cardinaliteiten. Deze Bell ongelijkheden worden op hun beurt herschreven in termen van vage, scalaire cardinaliteiten[1]. De kans dat twee gebeurtenissen samen optreden, wordt dan gemodelleerd aan de hand van een commutatieve conjunctor. Bovendien tonen we aan dat de Bell ongelijkheden voor commutatieve conjunctoren nodige en voldoende voorwaarden zijn zodanig dat de corresponderende Bell ongelijkheden voor vage probabiliteiten vervuld zijn voor elke vaagverzameling.

[1]In dit werk beschouwen we immers enkel de vage, scalaire cardinaliteit (de zogenaamde σ-count) als mogelijkheid voor de cardinaliteit van een vaagverzameling.

We richten onze aandacht echter niet alleen op commutatieve conjunctoren, maar ook op een aantal bekende klassen ervan, namelijk quasi-copulas, copulas en driehoeksnormen. We tonen aan dat sommige Bell ongelijkheden voldaan zijn voor (quasi-)copulas. Verder beschouwen we de Frank familie van driehoeksnormen en hebben we, voor deze familie, de waarden van de parameter bepaald zodat de Bell ongelijkheden vervuld zijn.

Ook het verband tussen de Bell ongelijkheden en continue driehoeksnormen komt in dit werk aan bod. Een belangrijk resultaat is dat ordinale sommen de Bell ongelijkheden bewaren zodat het voldoende is om enkel continue, Archimedische driehoeksnormen te bestuderen. De belangrijkste parametrische families van driehoeksnormen (die opgesomd worden door Klement, Mesiar en Pap) worden in dit werk onderzocht.

De resultaten die we verkrijgen voor de vervaagde Bell ongelijkheden kunnen nu gebruikt worden om twee belangrijke stellingen te formuleren die ervoor zorgen dat meer algemene vergelijkingen met betrekking tot vage cardinaliteiten gemakkelijk kunnen geverifieerd worden (met als gevolg dat herhaalde berekeningen vermeden worden). Deze twee stellingen worden als volgt opgedeeld: een eerste stelling heeft enkel betrekking op twee vaagverzamelingen, terwijl de tweede stelling drie vaagverzamelingen in acht neemt. Beide stellingen drukken echter hetzelfde uit: wanneer een ongelijkheid (met betrekking tot cardinaliteiten) geldt voor scherpe verzamelingen en de commutative conjunctor, die gebruikt wordt om de doorsnede van twee vaagverzamelingen te modelleren, voldoet aan bepaalde Bell ongelijkheden, dan geldt deze ongelijkheid ook voor vaagverzamelingen. Voor beide stellingen formuleren we eveneens een meer specifieke stelling voor ongelijkheden die niet afhangen van de cardinaliteit van het universum.

Vage similariteitsmaten worden reeds frequent toegepast in verschillende domeinen (bijvoorbeeld in de beeldverwerking, in vage modelleringstoepassingen of bij het extraheren van informatie uit documenten). Het is natuurlijk steeds mogelijk om met niets te beginnen en zo een vage similariteitsmaat op te bouwen, maar er bestaat een veel simpelere methode. We vertrekken hiervoor van een similariteitsmaat voor scherpe verzamelingen (of equivalent hiermee, binaire vectoren) en stellen een schema met vervagingsregels op om zo een vage similariteitsmaat te construeren. De bestaande toepassingen met vage similariteitsmaten zijn alle van deze aard. De meest gebruikte similariteitsmaat is een vervaging van de Jaccard coefficient, alhoewel vervagingen van de Dice coefficient, de Simple Matching coefficient of de Ochiai coefficient niet ontbreken. Bij de vervaging van deze similariteitsmaten wordt echter bijna altijd gebruik gemaakt van de minimum en maximum operator om respectievelijk doorsnede en unie van twee vaagverzamelingen te modelleren.

Het belangrijkste resultaat omtrent de geparametriseerde familie vage similariteitsmaten is dan ook dat de transitiviteitseigenschappen die vervuld zijn voor scherpe verzamelingen, ook geldig blijven voor vaagverzamelingen op voorwaarde dat de doorsnede van twee vaagverzamelingen gemodelleerd wordt aan de hand van een commutatieve quasi-copula.

Omtrent vage inclusiematen is veel meer te vinden in de literatuur dan hun tegenhangers in het scherpe geval. De meeste vage inclusiematen worden geconstrueerd aan de hand van een implicator (deze maten zijn gebaseerd op de definitie van scherpe inclusie: $A \subseteq B \Leftrightarrow (\forall x \in X)(x \in A \Rightarrow x \in B)$). Sommige auteurs verkiezen echter een axiomatische aanpak om een vage inclusiemaat te definiëren. Wij nemen echter opnieuw als vertrekpunt de familie inclusiematen voor scherpe verzamelingen om een familie van vage inclusiematen op te stellen. Op deze manier voorzien we de lezer van een heel scala aan vage inclusiematen. Bovendien modelleren we opnieuw de doorsnede van twee vaagverzamelingen aan de hand van een commutatieve quasi-copula zodat Łukasiewicz-transitiviteit ook na het vervagingsproces behouden blijft.

Tenslotte tonen we aan dat zowel vage similariteitsmaten als vage inclusiematen hun nut kunnen bewijzen in reële toepassingen.

Bibliography

[1] A. Alsina, R. Nelsen, and B. Schweizer, *On the characterization of a class of binary operations on distribution functions*, Statistics and Probability Letters **17** (1993), 85–89.

[2] W. Bandler and L. Kohout, *Fuzzy power sets and fuzzy implication operators*, Fuzzy Sets and Systems **4** (1980), 13–30.

[3] C. Baroni-Urbani and W. Buser, *Similarity of binary data*, Systematic Zoology **25** (1976), 251–259.

[4] J. S. Bell, *On the Einstein–Podolsky–Rosen paradox*, Physics **1** (1964), 195–200.

[5] N. Blanchard, *Cardinal and ordinal theories about fuzzy sets*, Fuzzy Information and Decision Processes (M.M. Gupta and E. Sanchez, eds.), North-Holland, Amsterdam, 1982, pp. 149–157.

[6] I.J. Blanco, M.J. Martín-Bautista, D. Sánchez, and M.A. Vila, *A new similarity measure for effectiveness in fuzzy information retrieval*, Proc. IPMU (Perugia, Italy), 2004, pp. 1991–1997.

[7] U. Bodenhofer, *A new approach to fuzzy orderings*, Tatra Mt. Math. Publ. (16) (1999), 21–29.

[8] D. Bohm, *Quantum theory*, Prentice-Hall, Englewood Cliffs, New York, 1951.

[9] S.A. Boorman and D.C. Olivier, *Metrics on spaces of finite trees*, Journal of Mathematical Psychology **10** (1973), 26–59.

[10] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel, *Towards general measures of comparison of objects*, Fuzzy Sets and Systems **84** (1996), 143–153.

[11] J. Braun-Blanquet, *Plant sociology: the study of plant communities*, translated by G.D. Fuller and H.S. Conrad ed., McGraw-Hill, New York, 1932.

[12] D. Bryant, J. Tsang, P. Kearney, and M. Li, *Computing the quartet distance between evolutionary trees*, Proc. Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, California), 2000, pp. 285–286.

[13] P. Burillo, N. Frago, and R. Fuentes, *Inclusion grade and fuzzy implication operators*, Fuzzy Sets and Systems **114** (2000), 417–429.

[14] J. Casasnovas and J. Torrens, *Scalar cardinalities of finite fuzzy sets for t-norms and t-conorms*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems **11** (2003), 599–614.

[15] T. Chaira and A.K. Ray, *Fuzzy measures for color image retrieval*, Fuzzy Sets and Systems **150** (2005), 545–560.

[16] M. Chen and D.A. Linkens, *Rule-base self-generation and simplification for data-driven fuzzy models*, Fuzzy Sets and Systems **142** (2004), 243–265.

[17] H.T. Clifford and W. Stephenson, *An Introduction to Numerical Classification*, Academic Press, London, 1975.

[18] B. Dasgupta, X. He, T. Jiang, M. Li, J. Tromp, and L. Zhang, *On distances between phylogenetic trees*, Proc. Eight Annual ACM-SIAM Symposium on Discrete Algorithms (New Orleans, Louisiana), 1997, pp. 427–436.

[19] B. De Baets and H. De Meyer, *Transitivity-preserving fuzzification schemes for cardinality-based similarity measures*, European J. Oper. Res. **160** (2005), 726–740.

[20] B. De Baets, H. De Meyer, and H. Naessens, *A class of rational cardinality-based similarity measures*, J. Comp. Appl. Math. **132** (2001), 51–69.

[21] B. De Baets, H. De Meyer, and H. Naessens, *On rational cardinality-based inclusion measures*, Fuzzy Sets and Systems **128** (2002), 169–183.

[22] B. De Baets, S. Janssens, and H. De Meyer, *Meta-theorems on inequalities for scalar fuzzy set cardinalities*, Fuzzy Sets and Systems **157** (2006), 1463–1476.

[23] B. De Baets and R. Mesiar, *Pseudo-metrics and T-equivalences*, J. Fuzzy Math. **5** (1997), 471–481.

[24] B. De Baets and R. Mesiar, *Ordinal sums of aggregation operators*, Technologies for Constructing Intelligent Systems: Tools (B. Bouchon-Meunier, J. Gutiérrez-Ríos, L. Magdalena, and R. Yager, eds.), Springer–Verlag, 2002, pp. 137–147.

[25] A. De Luca and S. Termini, *A definition of non-probabilistic entropy in the setting of fuzzy sets theory*, Information and Control **20** (1972), 301–312.

[26] H. De Meyer, B. De Baets, and S. Janssens, *Similarity measurement on leaf-labelled trees*, Proc. 2nd EUSFLAT Conference (Leicester, U.K.), 2001, pp. 253–256.

[27] M. Delgado, D. Sanchez, M.J. Martin-Bautista, and M.A. Vila, *A probabilistic definition of a nonconvex fuzzy cardinality*, Fuzzy Sets and Systems **126** (2002), 177–190.

[28] L.R. Dice, *Measures of the amount of ecologic associations between species*, Ecology **26** (1945), 297–302.

[29] D. Dubois and H. Prade, *Fuzzy sets and systems: Theory and applications*, Mathematics in Science and Engineering, Vol. 144, Academic Press, New York, 1980.

[30] D. Dubois and H. Prade, *Fuzzy cardinality and the modelling of imprecise quantification*, Fuzzy Sets and Systems **16** (1985), 199–230.

[31] G.F. Estabrook, F.R. McMorris, and C.A. Meacham, *Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units*, Systematic Zoology **2** (34) (1985), 193–200.

[32] E.W. Fager and J.A. McGowan, *Zooplankton species groups in the North Pacific*, Science **140** (1963), 453–460.

[33] J. Fan, W. Xie, and J. Pei, *Subsethood measure: new definitions*, Fuzzy Sets and Systems **106** (1999), 201–209.

[34] M.A. Fligner, J.S. Verducci, and P.E. Blower, *A modification of the Jaccard–Tanimoto similarity index for diverse selection of chemical compunds using binary strings*, Technometrics **44** (2002), 110–119.

[35] M. Frank, *On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$*, Aequationes Math. **19** (1979), 141–160.

[36] K. Fukuda and V. Rosa, *Combinatorial face enumeration in convex polytopes*, Computational Geometry **4** (1994), 191–198.

[37] C. Genest, L. Molina, L. Lallena, and C. Sempi, *A characterization of quasi-copulas*, Journal of Multivariate Analysis **69** (1999), 193–205.

[38] S. Gottwald, *A note on fuzzy cardinals*, Kybernetika **16** (1980), 156–158.

[39] J.C. Gower and P. Legendre, *Metric and euclidean properties of dissimilarity coefficients*, Journal of Classification **3** (1986), 5–48.

[40] J.W. Holley and J.P. Guilford, *A note on the g index of agreement*, Psychological Measurement **24** (1964), 749–753.

[41] P. Jaccard, *Nouvelles recherches sur la distribution florale*, Bulletin de la Sociétée Vaudoise des Sciences Naturelles **44** (1908), 223–270.

[42] S. Janssens, B. De Baets, and H. De Meyer, *On the transitivity of a parametrized family of similarity measures*, Proc. EUROFUSE Workshop on Information Systems (Varenna, Italy), 2002, pp. 239–243.

[43] S. Janssens, B. De Baets, and H. De Meyer, *Transitivity of comparison measures*, Proc. FUZZ-IEEE (Honolulu, Hawaï), 2002, (CD), pp. 1369–1374.

[44] S. Janssens, B. De Baets, and H. De Meyer, *Inequalities in fuzzy probability calculus*, Lecture Notes in Computer Science **2715** (2003), 211–218.

[45] S. Janssens, B. De Baets, and H. De Meyer, *Some meta-theorems on fuzzy cardinalities and their application*, Proc. 3rd EUSFLAT conference (Zittau, Germany), 2003, pp. 318–321.

[46] S. Janssens, B. De Baets, and H. De Meyer, *Identities and inequalities on fuzzy set cardinalities*, Proc. of NAFIPS (Montréal, Canada), 2006, (CD, 6 pages).

[47] S. Janssens, B. De Baets, and H. De Meyer, *Bell-type inequalities for parametric families of triangular norms*, Kybernetika **40** (2004), 89–106.

[48] S. Janssens, B. De Baets, and H. De Meyer, *Bell-type inequalities for quasi-copulas*, Fuzzy Sets and Systems **148** (2004), 263–278.

[49] A. Kaufmann, *Introduction à la Théorie des Sous-Ensembles Flous*, Masson, Paris, 1977.

[50] P. Kilpeläinen and H. Mannila, *Ordered and unordered tree inclusion*, SIAM Journal of Computing **24** (1995), 340–356.

[51] L. Kitainik, *Fuzzy Decision Procedures with Binary Relations*, Kluwer Academic Publishers, Boston, 1993.

[52] E.P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*, Trends in Logic, Studia Logica Library, Vol. 8, Kluwer, Dordrecht, 2000.

[53] E.P. Klement, R. Mesiar, and E. Pap, *Invariant copulas*, Kybernetika **38** (2002), 275–285.

[54] A. Kolesárová, *1-Lipschitz aggregation operators and quasi-copulas*, Kybernetika **39** (2003), 615–629.

[55] B. Kosko, *Neural Networks and Fuzzy Systems: a Dynamic Systems Approach to Machine Intelligence*, Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[56] S. Kulczynski, *Die pflanzenassoziationen der pieninen*, Bulletin International de l'Académie Polonaise des Sciences et des Lettres (1927), 57–203.

[57] L.I. Kuncheva, *Fuzzy rough sets: application to feature selection*, Fuzzy sets and systems **51** (1992), 147–153.

[58] C. M. Ling, *Representation of associative functions*, Publ. Math. Debrecen **12** (1965), 189–212.

[59] B.G. Mirkin, *Clustering for Data Mining: a Data Recovering Approach*, Computer Science and Data Analysis Series, Vol. 3, Chapman & Hall/CRC, 2005.

[60] T. Motzkin, H. Raiffa, G. Thompson, and R. Thrall, *The double description method*, Contributions to Theory and Games (H. Kuhn and A. Tucker, eds.), Vol. 2, Princeton University Press, Princeton, RI, 1953.

[61] R. Nelsen, *An Introduction to Copulas*, Lecture Notes in Statistics, Vol. 139, Springer–Verlag, New York, 1999.

[62] A. Ochiai, *Zoogeographic studies on the soleoid fishes found in japan and its neighbouring regions*, Bull. Jpn. Soc. Sci. Fish. **22** (1957), 526–530.

[63] J.F. Omhover, M. Detyniecki, and B. Bouchon-Meunier, *A region-similarity-based image retrieval system*, Proc. IPMU (Perugia, Italy), 2004, pp. 1461–1468.

[64] C. Pappis and N. Karacapilidis, *A comparative assessment of measures of similarity of fuzzy values*, Fuzzy Sets and Systems **56** (1993), 171–174.

[65] D. Penny and M.D. Hendy, *The use of tree comparison metrics*, Systematic Zoology **1** (34) (1985), 75–82.

[66] I. Pitowsky, *Quantum Probability – Quantum Logic*, Lecture Notes in Physics, Vol. 321, Springer, Berlin, New York, 1989.

[67] J. Pykacz and B. D'Hooghe, *Bell-type inequalities in fuzzy probability calculus*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems **9** (2001), 263–275.

[68] D. Ralescu, *Cardinality, quantifiers and the aggregation of fuzzy criteria*, Fuzzy Sets and Systems **69** (1995), 355–365.

[69] D. Rogers and T. Tanimoto, *A computer program for classifying plants*, Science **132** (1960), 1115–1118.

[70] P.F. Russell and T.R. Rao, *On habitat and association of species of anopheline larvae in south-eastern madras*, Journal of the Malaria Institute of India **3** (1940), 153–178.

[71] S. Santini and R. Jain, *Similarity measures*, IEEE Trans. Pattern Anal. Mach. Intell. **21** (1999), 871–883.

[72] B. Sarker, *The resemblance coefficients in group technology: a survey and comparative study of relational metrics*, Computers ind. Engng **1** (1996), 103–116.

[73] S.M. Selkow, *The tree-to-tree editing problem*, Information Processing Letters **6** (1977), 184–186.

[74] R.P. Sheridan and S.K. Kearsley, *Why do we need so many chemical similarity search methods?*, Drug Discovery Today **7** (2002), 903–911.

[75] G.G. Simpson, *Notes on the measurement of faunal resemblance*, American Journal of Science **258** (1960), 300–311.

[76] D. Sinha and E.R. Dougherty, *Fuzzification of set inclusion: theory and applications*, Fuzzy Sets and Systems **55** (1993), 15–42.

[77] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8** (1959), 229–231.

[78] R. Sokal and C. Michener, *A statistical method for evaluating systematic relationships*, Univ. of Kansas Science Bulletin **38** (1958), 1409–1438.

[79] R. Sokal and P. Sneath, *Principles of Numerical Taxonomy*, W.H. Freeman, San Francisco, 1963.

[80] Y. Tolias, S. Panas, and L. Tsoukalas, *Generalized fuzzy indices for similarity matching*, Fuzzy Sets and Systems **120** (2001), 255–270.

[81] A. Tversky, *Features of similarity*, Psychological Review **84** (1977), 327–352.

[82] C.J. van Rijsbergen, *Information Retrieval*, Butterworths, London, 1979.

[83] J.T.L. Wang, H. Shan, D. Shasha, and W.H. Piel, *Treerank: a similarity measure for nearest neighbour searching in phylogenetic databases*, Proc. 15th Intern. Conf. on Scientific and Statistical Database Management (Cambridge, Massachusetts), 2003, pp. 171–180.

[84] X. Wang, B. De Baets, and E. Kerre, *A comparative study of similarity measures*, Fuzzy Sets and Systems **73** (1995), 259–268.

[85] S.M. Waterman and T.F. Smith, *On the similarity of dendrograms*, Journal of Theoretical Biology **73** (1978), 789–800.

[86] P. Willett, *Recent trends in hierarchic document clustering*, Information Processing and Management **24** (1988), 577–597.

[87] P. Willett, *Chemical similarity searching*, J. Chem. Inf. Comput. Sci. **38** (1998), 983–996.

[88] P. Willett, *Similarity-based approaches to virtual screening*, Biochemical Society Transactions **31** (2003), 603–606.

[89] R. Willmott, *Two fuzzier implication operators in the theory of fuzzy power sets*, Fuzzy Sets and Systems **4** (1980), 31–36.

[90] R. Willmott, *On the transitivity of containment and equivalence in fuzzy power set theory*, J. Math. Anal. Appl. **120** (1986), 384–396.

[91] M. Wygralak, *Fuzzy cardinals based on the generalized equality of fuzzy subsets*, Fuzzy Sets and Systems **18** (1986), 338–353.

[92] M. Wygralak, *An axiomatic approach to scalar cardinalities of fuzzy sets*, Fuzzy Sets and Systems **110** (2000), 175–179.

[93] V.R. Young, *Fuzzy subsethood*, Fuzzy Sets and Systems **77** (1996), 371–384.

[94] G.U. Yule and M.G. Kendall, *An Introduction to the Theory of Statistics*, 14 ed., Hafner, 1950.

[95] L.A. Zadeh, *Fuzzy sets*, Information and Control **8** (1965), 338–353.

[96] L.A. Zadeh, *Similarity relations and fuzzy orderings*, Information Sciences **3** (1971), 177–200.

[97] L.A. Zadeh, *A theory of approximate reasoning*, Machine Intelligence **9** (1979), 149–194.

[98] L.A. Zadeh, *A computational approach to fuzzy quantifiers in natural languages*, Comput. Math. Appl. **9** (1983), 149–184.

[99] L.A. Zadeh, *Fuzzy probabilities*, Information Processing and Management **20** (1984), 363–372.

[100] Y. Zhong, C.A. Meacham, and S. Pramanik, *A general method for tree-comparison based on subtree similarity and its use in a taxonomic database*, BioSystems **42** (1997), 1–8.