LINZ 2007

28^{th} Linz Seminar on Fuzzy Set Theory

Abstracts

Fuzzy Sets, Probability, and Statistics

—

Gaps and Bridges

Bildungszentrum St. Magdalena, Linz, Austria
February 6–10, 2007

Abstracts

Didier Dubois, Erich Peter Klement, Radko Mesiar
Editors

LINZ 2007

—

FUZZY SETS, PROBABILITY, AND STATISTICS – GAPS AND BRIDGES

ABSTRACTS

Didier Dubois, Erich Peter Klement, Radko Mesiar
Editors

Printed by: Universitätsdirektion, Johannes Kepler Universität, A-4040 Linz

Since their inception in 1979 the Linz Seminars on Fuzzy Sets have emphasized the development of mathematical aspects of fuzzy sets by bringing together researchers in fuzzy sets and established mathematicians whose work outside the fuzzy setting can provide direction for further research. The seminar is deliberately kept small and intimate so that informal critical discussion remains central. There are no parallel sessions and during the week there are several round tables to discuss open problems and promising directions for further work.

LINZ 2007, already the 28th seminar carrying on this tradition, will be devoted to the mathematical aspects of "Fuzzy Sets, Probability, and Statistics – Gaps and Bridges". As usual, the aim of the Seminar is an intermediate and interactive exchange of surveys and recent results.

Didier Dubois
Erich Peter Klement
Radko Mesiar

Program Committee

Didier Dubois (Chairman), *Toulouse, France*
Radko Mesiar, *Bratislava, Slovak Republic*
Ulrich Bodenhofer, *Hagenberg, Austria*
Dan Butnariu, *Haifa, Israel*
Bernard De Baets, *Gent, Belgium*
János Fodor, *Budapest, Hungary*
Lluis Godo, *Barcelona, Spain*
Siegfried Gottwald , *Leipzig, Germany*
Petr Hájek, *Praha, Czech Republic*
Ulrich Höhle, *Wuppertal, Germany*
Erich Peter Klement, *Linz, Austria*
Wesley Kotzé, *Cape Town, South Africa*
Daniele Mundici, *Firenze, Italy*
Endre Pap, *Novi Sad, Serbia and Montenegro*
Stephen E. Rodabaugh, *Youngstown, OH, USA*
Lawrence N. Stout, *Bloomington, IL, USA*
Aldo Ventre, *Napoli, Italy*
Siegfried Weber, *Mainz, Germany*

Executive Committee

Erich Peter Klement
Ulrich Höhle
Stephen E. Rodabaugh
Siegfried Weber

Local Organizing Committee

Erich Peter Klement (Chairman), *Johannes Kepler University Linz*
Ulrich Bodenhofer, *Johannes Kepler University Linz*
Fabrizio Durante, *Johannes Kepler University Linz*
Sandra Haghofer, *Software Competence Center Hagenberg*
Sabine Lumpi, *Fuzzy Logic Laboratorium Linz-Hagenberg*
Bernhard Moser, *Software Competence Center Hagenberg*
Susanne Saminger, *Johannes Kepler University Linz*
Peter Sarkoci, *Johannes Kepler University Linz*
Isabel Tober-Kastner, *Software Competence Center Hagenberg*
Gerolf Seitz, *Software Competence Center Hagenberg*

Contents

Quasi-Concave Copulas, Asymmetry, Transformations

Elisabetta Alvoni[1] and Pier Luigi Papini[2]

[1] Dipartimento MATEMATES, Università di Bologna,
Viale Filopanti 5, I-40126 Bologna, Italy
elisabetta.alvoni@unibo.it
[2] Dipartimento di Matematica, Università di Bologna,
Piazza Porta S.Donato 5, I-40126 Bologna, Italy
papini@dm.unibo.it

In this paper we consider a class of copulas, called quasi-concave; we compare them with other classes of copulas and we study conditions implying symmetry for them.

Recently, a measure of asymmetry for copulas has been introduced and the maximum degree of asymmetry for them in this sense has been computed: see [5, 3].

Here we compute the maximum degree of asymmetry that quasi-convex copulas can have; we prove that the supremum of

$$\{|C(x,y) - C(y,x)|; x, y \in [0,1]; C \text{ is quasi-concave}\}$$

is $1/5$.

Also, we show by suitable examples that such supremum is a maximum and we indicate copulas for which the maximum is achieved.

Moreover, we show that the class of quasi-concave copulas is preserved by a class of simple transformations, often considered in the literature.

References

1. F. Durante, Solution of an open problem for associative copulas, Fuzzy Sets Syst. 152 (2005), 411-415.
2. C. Genest, K. Ghoudi and L.-P. Rivest, Discussion on "Understanding relationships using copulas" by E. Frees and E. Valdez; North American Actuarial Journal, 2 (1999), 143-149.
3. E.P. Klement and R. Mesiar, How non-symmetric can a copula be? Comment. Math. Univ. Carolinae 47 (2006), 141-148.
4. E.P. Klement, R. Mesiar and E. Pap, Different types of continuity of triangular norms revisited, New Math. and Natural Comput. 1 (2005), 1-17.
5. R.B. Nelsen, Extremes of nonexchangeability, Statist. Papers, to appear.
6. R.B. Nelsen, An introduction to copulas, 2nd edition; Springer, New York, 2006.

Maximum Possibility vs. Maximum Likelihood Decisions

Luca Bortolussi[1], Massimo Borelli[1], and Andrea Sgarro[1,2]

[1] Dept. of Mathematics and Computer Science,
University of Trieste,
34100 Trieste, Italia
luca@dmi.units.it, {borelli|sgarro}@units.it
[2] Consorzio di Biomedicina Molecolare,
Area Science Park, 34100 Trieste.

Consider the following very familiar decision-theoretic situation: a list $\mathcal{L}$ is chosen out of a (finite) input set $\mathcal{X}$, and is communicated to an observer. Further, an input object x, sometimes called a *state of nature*, is chosen inside $\mathcal{L}$. The observer cannot observe directly x, but only a "corrupted version" thereof, y say. He/she makes the following decision: decide for the objects $d(y)$ in the list $\mathcal{L}$ which are "most similar" to what he/she could observe, i.e. to y. Clearly, all this assumes that *similarity measures* $\sigma(x,y)$ are given between input and output objects (between the states of nature and the observables): we shall arrange these measures into a *similarity matrix* Σ with rows headed to $\mathcal{X}$ and columns headed to the (finite) output set $\mathcal{Y}$. The entries of Σ are non-negative real numbers; to avoid trivial situations, at least one entry is strictly positive.

In a coding-theoretic approach, as pursued in [3], the list $\mathcal{L}$ is called the *codebook*, and x and y are the *input codeword* and the *output word*, respectively[3]; then the similarity matrix would describe the *noise* which affects the communication channel. It is in coding theory, and more precisely in possibilistic coding theory and its application to DNA word design [1], that the motivation for this work[4] resides.

Some cases of special matrices follow, which fit into this general frame:

- *Stochastic matrix*: the sum of each row is equal to 1.
- *Joint probability matrix*: the sum of all entries is 1.
- *Possibilistic transition matrix*, or simply *possibility matrix*: the maximum entry in each row is 1.
- *Joint possibility matrix*: the maximum entry in the matrix is 1.

When Σ is a stochastic matrix (then similarities are conditional probabilities), the decision-theoretic principle above is simply *maximum likelihood*, while it is the bayesian principle of *maximum posterior probability* in the case of joint probabilities. As for possibility matrices, whose entries are transition possibilities (conditional possibilities), the reader is referred to [4] which deals with a coding-theoretic frame. One may envisage also a "bayesian" possibilistic case, with matrices of joint possibilities whose overall maximum is 1: in this case, each matrix entry is a joint possibility obtained by taking the minimum of the "prior" possibility of the input and the conditional possibility of the output given that input; cf. [2] where the underlying notion of *interactivity* is illustrated.

Assume that Σ is altered to Σ' without changing the orderings between entries. Operationally, nothing would change from the point of view of the decision $d(y)$ made by the observer, whatever

[3] In coding theory, when $|d(y)| \geq 2$ the decoder may either try to guess a single codeword inside $d(y)$, and by so doing increase the probability/possibility of an *undetected error*, or keep $d(y)$ as it is and declare a *detected error*.

[4] What we need in [1] is a communication model which is as unassuming as possible, as we are interested in "negative" results of the type: no noisy channel exists which would justify such and such combinatorial DNA code construction. Since in the sequel we shall concentrate on "singleton events" (elementary events, individual words), rather than compound events (sets of words), we do not even have to specify how we should "aggregate" similarities to obtain similarities between sets of input words and sets of output words. Cf. also the remarks on compound events which conclude this extended abstract.

the list $\mathcal{L}$, whatever the input object $x \in \mathcal{L}$, and whatever the output object observed y. We shall say in such a case that Σ and Σ' are *equivalent*; an obvious and "limit" case of equivalence is when the two matrices are *proportional*. We shall investigate properties which are *stable* with respect to equivalences. We stress that equivalence concerns only singletons (elementary events) and not an algebra of sets (of compound events).

A problem arises, that of comparing the *representational capacity* or *expressive power* of these approaches, in the sense that on may or may not find equivalent matrices. By just fitting in the suitable proportionality constant, one can prove the following obvious facts: a criterion for a similarity matrix to be equivalent to a possibility matrix is that the maximum similarity in each row is the same; a sufficient but not necessary condition for a similarity matrix to be equivalent to a stochastic matrix is that each row of the similarity matrix sums to the same number. However trivial, we shall stress these facts in the theorem below; in particular, the theorem explains why in the sequel we shall forget about joint probabilities or joint possibilities, and stick instead to similarities: whenever one deals with a similarity matrix, one may well as well think that one is dealing with joint possibilities or joint probabilities, after fitting in the suitable proportionality constant. (Exhibiting possibility matrices which cannot be simulated by means of equivalent stochastic matrices is quite easy; in the lemma below we state a necessary condition.)

Theorem 1. *The representational capacity of similarities, joint possibilities and joint probabilities is the same. The representational capacity of conditional probabilities and transition possibilities are incomparable; both are strictly less than the representational capacity of similarities.*

One may have "odd" similarity matrices, indeed. For example the minimum in row a might be strictly greater than the maximum in row b, which would make the input object b totally "useless". In the sequel, we shall add constraints to the definition of similarity matrices, so as to get rid of "strange" situations, and check how all this shrinks the corresponding representational capacity.

Certain input objects (codewords, states of nature) in a similarity matrix may be "redundant" in the sense of row domination: row a is *dominated* by row b when $a_i \leq b_i$. General similarity matrices or even possibility matrices may freely have domination between their rows, while stochastic matrices have it only in a limit case, since they verify the obvious property: if row a is dominated by row b, then $a = b$. Actually, stochastic matrices verify a stronger ordinal property, which involves domination for rows after *re-ordering* the row entries: in two rows of a similarity matrix there is an *inversion* when, after re-ordering the rows with respect to the non-decreasing order, say, there are two positions i and j with $a_i < b_i$, while $a_j > b_j$. Now, two rows exhibit no inversion iff a permutation of one of the two is dominated by the other.

Lemma 1. *For a similarity matrix to be equivalent to a stochastic matrix, there must be at least one inversion in each couple of rows, apart from couples of rows which are equal up to a permutation of their entries. This condition is also sufficient for two-row matrices.*

(Proof omitted in this extended abstract.) When a matrix satisfies the condition as in the lemma, for convenience' sake we shall say that the matrix is *regular*; we stress that regularity is a topological property which is *stable* with respect to matrix equivalence. The following three-line counter-example shows that this condition is *not* sufficient to have stochasticity up to an equivalence. Take the three-row similarity matrix

$$\begin{matrix} a & a & d & d \\ b & c & c & c \\ a & c & c & d \end{matrix}$$

with $a < b < c < d$; the three rows are already properly ordered. In rows 1 and 2 there is an inversion in positions (columns) 1 and 3, in rows 1 and 3 there is an inversion in positions 2 and 3, while in rows 2 and 3 there is an inversion in positions 1 and 4. However, the linear programming problem which one has to solve (details omitted in this extended abstract) is

$$a < b < c < d \ , \ \ 2a + 2d = 1 \ , \ \ b + 3c = 1 \ , \ \ a + 2c + d = 1 \ ,$$

whose solution set is empty: actually, the last two equations (after replacing $a+d$ by 1/2, cf. the first equation) give $b = c = 1/4$, while one should have $b < c$. Assuming $d < 1$ and adding an all-1 column shows that one can as well start from a possibility matrix.

Theorem 2. *The representational capacity of regular similarities (and so of regular joint possibilities) strictly exceeds that of stochastic matrices. The representational capacity of regular transition possibilities and that of stochastic matrices are not comparable; however, they are the same for two-row matrices.*

All this leaves open the following *open problem*, at least when the number of states of nature is at least 3: *find a simple criterion to ensure that a similarity matrix is equivalent to a stochastic matrix*. Unfortunately, at this point we are only able to provide a sufficient condition which ensures the equivalence, based on a suitable "geometry" of inversions, as will be given in the final version.

Compound events. If one moves from singletons (individual words) to compound events (sets of words), one would have to specify a suitable *aggregator*, which is the sum in the case of probabilities and the maximum for possibilities, and would presumably be an "abstract" aggregator in the general case of similarities. By the way, restricting ourselves to singletons, as we do below, makes it difficult to re-cycle classical results on qualitative probabilities [3], which e.g. require that the intersection of conditioning events is not void, unlike what happens when intersecting distinct singletons. Considering only singletons (elementary events, be they states of nature or codewords), is of no consequence as far as *decoding* (decision making) is concerned, since this depends only on how similarities are ordered in the similarity matrix; however, it does matter when it comes to evaluate the *error* that the decoder might make, which is an additive error of the form $\mathrm{Prob}(E|x)$ in the case of probabilities and a maxitive error of the form $\mathrm{Poss}(E|x)$ in the case of possibilities, with E made up of several[5] output objects (more general aggregators might be used to evaluate the error in the case of similarities). In other words, our concern here is only how decisions are *made*, and not also how decisions should be *evaluated*. If one wants a notion of equivalence such as to be significant also for error evaluation, one should require that the ordering is preserved also for compound events. This is definitely more assuming than above; e.g., it is quite easy to give two-rows examples where an inversion is not enough to have equivalence in this strong sense between a possibilistic and a stochastic matrix. Take the joint possibilities

$$\begin{matrix} a & b & b & c \\ a & a & d & d \end{matrix}$$

with $0 < a < b < c < d = 1$; there is an inversion e.g. in columns 2 and 3. ¿From the second row one has $2a + 2d = 1$, and so $b < d < 1/2$; instead, from the first row one has $a + 2b + c = 1$ and so, after subtracting b, $a + b + c = \mathrm{Prob}(a,b,c) > 1/2 > d$, while $\max(a,b,c) = \mathrm{Poss}(a,b,c) < d$. Add an all-1 column if you want to start from a possibilistic matrix.

[5] By the way, it is a moot point how to define terms of the form $\mathrm{Poss}(y|E)$ in the case of possibilities, let alone in the "abstract" case of similarities.

References

1. L. Bortolussi and A. Sgarro. Possibilistic channels for dna word design. In *Proceedings of SMPS 2006*, 2006.
2. D. Dubois and H. Prade, editors. *Fundamentals of Fuzzy Sets*. Kluwer, 2000.
3. T. L. Fine. *Theories of Probability*. Academic Press, 1973.
4. A. Sgarro. Possibilistic information theory: a coding-theoretic approach. *Fuzzy Sets and Systems*, 132:11–32, 2002.

Perception-Based Information in a Coherent Conditional Probability Setting

Giulianella Coletti[1] and Romano Scozzafava[2]

[1] Dip. Matematica e Informatica, Università di Perugia, Via Vanvitelli 1, 06123 Perugia (Italy)
coletti@dipmat.unipg.it

[2] Dip. Metodi e Modelli Matematici, Università di Roma "La Sapienza", via Scarpa 16, 00161 Roma (Italy)
romscozz@dmmm.uniroma1.it

Many papers, starting with the pioneering one by Zadeh (1965), have been devoted – during past years – to support the negative view maintaining that probability is inadequate to capture what is usually treated by fuzzy theory. This view is still supported by Zadeh in recent years: in particular we mention the paper Zadeh (2002), which contains also a number of relevant references. His thesis is that PT – standard Probability Theory – is not fit to offer solutions for many simple problems in which a key role is played by a "perception–based information".

We agree with Zadeh's position, inasmuch he specifies that by PT he means *"standard probability theory of the kind found in textbooks and taught in courses"*. Actually, many traditional aspects of probability theory are not so essential as they are usually considered; for example, the requirement that the set of all possible "outcomes" should be endowed with a beforehand given algebraic structure – such as a Boolean algebra or σ–algebra – or the aim at getting, for these outcomes, *uniqueness* of their probability values, with the ensuing introduction of suitable relevant assumptions – such as σ-additivity, maximum entropy, conditional independence, ... – or interpretations, such as a strict frequentist one, which unnecessarily restricts the domain of applicability.

In the approach to probability expounded in a series of papers – for the relevant references see the book (Coletti and Scozzafava, 2002) – the leading tool is that of *coherence*, a concept that goes back to de Finetti (1949, 1970) and which allows to handle also those situations where we need to assess a (conditional) probability P on an *arbitrary* set of (conditional) events.

Our starting point is a synthesis of the available information, expressed by one or more *events*: to this purpose, the concept of event must be given its more general meaning, *i.e. it must not be looked on just as a possible outcome* – a subset of the so–called "sample space", as it is usually done in PT – but expressed by a *proposition*. Moreover, events play a two–fold role, since we must consider not only those events which are the direct object of study, but also those which represent the relevant "state of information": in fact a bunch of *conditional* events, together with a relevant "partial" assessment of *conditional* probability, are the tools that allow to manage specific – conditional – situations and to update degrees of belief on the basis of the evidence. The role of coherence is in fact that of ruling this extension process; a similar theory – but only for *unconditional* events – is the probabilistic logic by N.J. Nilsson (1986), which is just a re-phrasing – with different terminology – of de Finetti's theory, as Nilsson (1993) himself acknowledges.

Let φ_X be any *property* – in the sequel, to simplify notation we will write simply φ in place of φ_X – related to the quantity X : notice that a *property*, even if expressed by a statement, does not single–out an *event*, since the latter needs to be expressed by a *nonambiguous* proposition that can be either *true* or *false*.

Consider now the **event** E_φ = "You claim φ" and a coherent conditional probability $P(E_\varphi|A_x)$, looked on as a real function $\mu_\varphi(x) = P(E_\varphi|A_x)$ defined on C_X.

Since the events A_x are incompatible, then every $\mu_\varphi(x)$ with values in $[0,1]$ is a coherent conditional probability. So we can *define* a fuzzy subset in the following way.

Definition 1 – *Given a quantity X with range C_X and a related property φ, a fuzzy subset E^*_φ of C_X is the pair*

$$E^*_\varphi = \{E_\varphi\,, \mu_\varphi\},$$

with $\mu_\varphi(x) = P(E_\varphi|A_x)$ for every $x \in C_X$.

Notice that this conditional probability $P(E_\varphi|A_x)$ is *directly* introduced as a function on the set of conditional events, and without assuming any given algebraic structure. Is that possible? In the usual (Kolmorogovian) approach to conditional probability the answer is NO, since the introduction of $P(E_\varphi|A_x)$ would require the consideration and the assessment of $P(E_\varphi \wedge A_x)$ and $P(A_x)$, assuming positivity of the latter. But this *could not be* a simple task: in fact in this context the only sensible procedure is to assign directly $P(E_\varphi|A_x)$. For example, it is possible to assign the conditional probability that "You claim this number is small" knowing its value x, but not necessarily the probability that "The number has the value x"; not to mention that, for many choices of the quantity X, the corresponding probability can be zero. These problems are easily by–passed in our framework.

In fact, due to the *direct* assignment of $P(E_\varphi|A_x)$ as a whole, the knowledge – or the assessment – of the "joint" and "marginal" unconditional probabilities $P(E_\varphi \wedge A_x)$ and $P(A_x)$ is not required; moreover, the *conditioning* event A_x – which *must* be a *possible* event – may have *zero probability*. So, conditioning in a coherent setting gives rise to a general scenario that makes the classic Radon–Nikodym procedure – and the relevant concept of *regularity* – neither necessary nor significant. This has been repeatedly discussed elsewhere: see, e.g., Coletti and Scozzafava (2005).

So a coherent conditional probability $P(E_\varphi|A_x)$ is clearly a measure of how much You, given the event $A_x = \{X = x\}$, are willing to *claim* the property φ, and it plays the role of the membership function of the fuzzy subset E^*_φ.

Notice also that the significance of the conditional event $E_\varphi|A_x$ is reinforced by looking on it as "a whole", avoiding a separate consideration of the two propositions E_φ and A_x.

A fuzzy subset E^*_φ is a *crisp set* when the *only* coherent assessment $\mu_\varphi(x) = P(E_\varphi|A_x)$ has range $\{0,1\}$, i.e. when the property φ is such that, for every $x \in C_X$, one has either $E_\varphi \wedge A_x = \emptyset$ or $A_x \subseteq E_\varphi$.

Remark 1 – *Let us emphasize that in our context the concept of* fuzzy event*, as introduced by Zadeh (1968), is nothing else than a proposition, i.e., an ordinary event, of the kind "You claim the property φ". So, according to the rules of conditional probability – in particular, the "disintegration" formula, often called in the relevant literature "theorem of total probability" – we can easily compute its probability as*

$$P(E_\varphi) = \sum_x P(A_x)P(E_\varphi|A_x) = \sum_x P(A_x)\mu_\varphi(x) \;,$$

which coincides with Zadeh's definition *of the probability of (what he calls) a "fuzzy" event.*

Notice that this result is only a *trivial consequence of probability rules* and *not* a definition, and it puts also under the right perspective the subjective nature of a membership function, showing once again that our approach to probability goes beyond – both syntactically and semantically – the traditional one, denoted PT by Zadeh.

In conclusion, our theory is *not* a *probabilistic* – in the usual traditional sense – interpretation of *fuzziness*, since a *conditional* probability is *not* a probability, except in the trivial case in which the conditioning event is fixed and we let the first one vary: notice that we are proceeding the other way round!.

For the formal definitions concerning fuzzy sets, we recall that our results are expounded in a series of papers (Coletti and Scozzafava, 1999, 2001, 2004), where we show not only how to define fuzzy subsets, but we also introduce in a very natural way the counterparts of the basic continuous T-norms and the corresponding dual T-conorms, bound to the former by *coherence*. In particular, all Frank's norms and conorms – see Frank (1979) – are captured in our framework.

Concerning problems in which a role is played by a perception–based information (to use Zadeh's terminology), we are not going to discuss any single example, but we refer to the *model constituted by some balls in a box*, in different situations concerning their number, size, color, and any other feature of interest, as discussed in (Coletti and Scozzafava, 2006). The radical thesis placed by Zadeh on the table is that PT has serious limitations that cannot be overcome within the conceptual structure of bivalent logic, and so it cannot provide tools for operating on perception–based information. Our thesis is that we are able to manage situations of this kind, since PT is just a trivial particular case of our general approach to probability – and, mainly, to *conditional* probability – through coherence, and this approach encompasses all the tools that are necessary to deal with the kind of problems raised by Zadeh.

Consider a box containing n balls of various sizes $s_1,...,s_k$ ($k \leq n$), with respective fractions $f_1,...,f_k$, and an experiment consisting in drawing a ball from the box. Let E_L be the event (referred to the drawn ball) *"You claim (the size is) large"*. The question is: what is the probability of E_L?

In our context the problem is trivial, since it amounts to the computation of the probability of the ("fuzzy") event E_L. Introduce the random variable $\mathcal{S}$, with range $\{s_1,...,s_k\}$, of the sizes of the balls in the box, and consider the conditional events $E_L|S_i$, with $S_i = \{\mathcal{S} = s_i\}$. So by a trivial computation we have:

$$P(E_L) = \sum_i P(E_L|S_i)P(S_i) = \sum_i \mu_L(s_i) f_i \,,$$

where μ_L is the membership function of the fuzzy set "large", which is context dependent, since it refers to the sizes of the balls in the given box. The same procedure can be obviously followed for the properties "medium" and "small".

To simplify the exposition, we may hallmark a "fuzzy" event by the symbol F followed by the relevant property. Recall that in our context a fuzzy event is a ... "normal" event, i.e., a *particular* proposition beginning with "You claim ...": for example, the event E_L = "You claim *large*" can be simply written as

$$E_L = F\text{-large},$$

and the range of the property "large" should be clear from the context (e.g., in this case, the sizes of the balls).

Consider now the case that the balls have different colors – with known fractions – and introduce the corresponding random variable $\mathcal{C}$, with range $c_1,...,c_m$. We can consider the fuzzy subset of $\mathcal{C}$ singled–out by the membership function

$$\mu_D(c_r) = P(E_D|C_r) \,,$$

where E_D = F-dark and $C_r = \{\mathcal{C} = c_r\}$. Then it is clearly possible to evaluate $P(E_D)$ by the same procedure followed above for $P(E_L)$.

If we want to refer to *both* properties – a ball that is large and dark – we need considering the event $E_L \wedge E_D$ = F-(large$\wedge$dark) , getting

$$P(E_L \wedge E_D) = \sum_{i,j} P\big((E_L \wedge E_D)\big|(S_i \wedge C_j)\big) p_{ij} = \sum_{i,j} \mu_{LD}(s_i, c_j) p_{ij} \,,$$

where p_{ij} denotes the probability of a size s_i and a color c_j: the values p_{ij} can be computed by disintegration with respect to the possible compositions of the box, possibly taking into account suitable conditional independence assumptions. Concerning the membership $\mu_{LD}(s_i, c_j)$, any value belonging to the interval of coherence does the job (and also any Frank t-norm: in particular, a possible choice is the usual "minimum", as done by Zadeh)

Suppose now that we ignore the color of the balls and that we do not know the fractions f_i of balls having the various sizes s_i, *i.e.*, the box is of *unknown composition*. Consider the event, referred to the balls in the box

$E_{M,L}$="You claim that most are large"= F-(most are large).

The relevant claim may be seen as the assignment of a high probability of being claimed – referred to a ball to be drawn – large. Clearly, the possible values of this probability are identified with the possible values of the fractions p_i of large balls, and these values constitute the range of a random variable Π, related to the events $\Pi_i = \{\Pi = p_i\}$. Then we can introduce the membership function $\mu_M(p_i) = P(E_M|\Pi_i)$, with $E_M = F$-high, referred to the probability of drawing a large ball.

Going back to $E_{M,L}$, one of the questions posed by Zadeh is: assuming $E_{M,L}$, *i.e.* that most balls are large, what is the probability of drawing a large ball?

In our setting and notation, this amounts to the evaluation of the *conditional* probability $P(E_L|E_{M,L})$. Introducing events H_k $(k = 1, 2, \ldots, r)$ denoting the possible compositions of the box with respect to the possible sizes of the balls, we may resort to the disintegration formula

$$P(E_L|E_{M,L}) = \sum_k P(E_L|H_k \wedge E_{M,L})P(H_k|E_{M,L}) = \sum_k P(E_L|H_k)P(H_k|E_{M,L}),$$

the latter equality coming from E_L being *conditionally independent* of $E_{M,L}$ given H_k; in fact, for any *known* composition of the box, the probability of drawing a large ball does not depend on the knowledge that most balls are large.

Recalling the random variable Π introduced above, we can represent the probability $P(E_L|H_k)$ appearing in the previous formula as

$$P(E_L|H_k) = \sum_i P(E_L|\Pi_i \wedge H_k)P(\Pi_i|H_k) = \sum_i P(E_L|\Pi_i)P(\Pi_i|H_k),$$

while the probability $P(H_k|E_{M,L})$ can be easily computed via Bayes' theorem and $P(E_{M,L}|H_k)$.

References

1. Coletti, G., Scozzafava, R. (1999). Conditional Subjective Probability and Fuzzy Theory. Proc. of 18th NAFIPS International Conference, IEEE, New York, 77–80.
2. Coletti, G., Scozzafava, R. (2001). Fuzzy sets as conditional probabilities: which meaningful operations can be defined?. Proc. of 20th NAFIPS International Conference, IEEE, Vancouver, 1892–1895.
3. Coletti, G., Scozzafava, R. (2002). *Probabilistic Logic in a Coherent Setting*, Kluwer (Series *Trends in Logic*, n.15), Dordrecht.
4. Coletti, G., Scozzafava, R. (2004). Conditional probability, fuzzy sets and possibility: a unifying view. Fuzzy Sets and Systems 144, 227–249.
5. Coletti, G., Scozzafava, R. (2005). Conditioning in a coherent setting: theory and applications. Fuzzy Sets and Systems 155, 26–49.
6. Coletti, G., Scozzafava, R. (2006). Conditional probability and fuzzy information. Computational Statistics and Data Analysis 51, 115–132.
7. de Finetti, B. (1949). Sull'impostazione assiomatica del calcolo delle probabilità. Annali Univ. Trieste 19, 3–55. - *Engl. transl.* in: Ch.5 of *Probability, Induction, Statistics*, Wiley, London, 1972.
8. de Finetti, B. (1970). *Teoria della probabilità*, Einaudi, Torino. - *Engl. transl.*: "Theory of Probability", Voll. 1 and 2., Wiley, Chichester, 1974.
9. Frank, M.J. (1979). On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$. Aequationes Math. 19, 194–226.
10. Nilsson, N.J. (1986). Probabilistic Logic. Artificial Intelligence 28, 71–87.

11. Nilsson, N.J. (1993). Probabilistic Logic Revisited. Artificial Intelligence 59, 39–42.
12. Zadeh, L.A. (1965). Fuzzy sets. Information and Control 8, 338–353.
13. Zadeh, L.A. (1968). Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications 23, 421–427.
14. Zadeh, L.A. (2002). Toward a perception–based theory of probabilistic reason- ing with imprecise probabilities. Journal of Statistical Planning and Inference 105, 233–264.

Conditional Possibility: From Numerical to Qualitative Approach

Giulianella Coletti[1] and Barbara Vantaggi[2]

[1] Dept. Matematica e Informatica,
Università di Perugia, Italy
coletti@dipmat.unipg.it

[2] Dept. Metodi e Modelli Matematici,
Università "La Sapienza", Rome, Italy
vantaggi@dmmm.uniroma1.it

1 Introduction

Any assessment of an (unconditional) uncertainty measure on a set of events $\mathcal{A}$ can be seen as the relevant degrees of belief measured by a scale "calibrated" for taking into account all the events at the same time; in other words any event is regarded as embedded on the macrocosm consisting of the complete family. Obviously in this macro-context it is possible that some differences among the degrees of belief are not captured through this scale (unconditional measure). On the contrary the difference of the degrees of belief of any two events will emerge when we consider the microcosm consisting of just the two events, and so by using a more sensible scale.

The model apt to manage this complex system, in which different hypotheses (or information) are taken into account simultaneously, is a conditional uncertainty measure, where we consider it in the most general way, that is as a primitive concept.

The main feature of such an approach resides in adopting a *direct* introduction of the conditional measure as a function whose domain is a set of conditional events $E|H$, so that it can be defined for any pair of events E, H in $\mathcal{A}$, with $H \neq \emptyset$.

Therefore in this context conditioning is not just a trivial modification of the "world". In fact, it is essential to regard conditioning events as "variables" or, in other words, as uncertain events which can be either true or false. This framework gives the opportunity to the decision maker (or the field expert) to take into account at the same time all the possible scenarios (represented by the relevant conditioning events).

Starting from probability [8] many models based on a *direct* definition of conditional uncertainty measures have been given in literature (see, for instance, [6, 1–4]).
Nevertheless the discussion about the best definition of conditional model is open and is a problem of long-standing interest in different research fields such as uncertainty reasoning, economic and decision models.

In this work we focus on possibility and necessity measures. We recall that in this setting various definitions of conditional possibilities have been introduced mainly by analogy with Kolmogorovian probabilistic framework. Hence, these definitions have in common the fact that a conditional measure is obtained as a derived concept from an "unconditional" one (see e.g. [7, 11, 14]). Starting from a possibility measure $\Pi(\cdot)$, a T-conditional possibility (with T a triangular norm) $\Pi(\cdot|H)$ is essentially defined for every E as a solution x of the equation $\Pi(E \wedge H) = T(x, \Pi(H))$. Obviously, this equation has not a unique solution for any $\Pi(E \wedge H)$ and $\Pi(H)$. Note that while for strictly increasing t-norms T to have a unique solution it is sufficient to discard events with zero possibility, this does not happen for other t-norms. For example, for $T = \min$ the above equation has not a unique solution for $\Pi(E \wedge H) = \Pi(H)$. Then, particular principles (as, e.g., minimum specificity [11]), which give rise to different definitions of conditional possibility, have been introduced. Essentially, this notion includes the aforementioned definitions: it gives more freedom.

Recently in [1] (as mentioned above) a general notion of conditional possibility has been introduced as a function on a suitable set of conditional events which satisfies a set of axioms. This definition removes all the critical cases and permits to consider $\Pi(E|H)$ for every pair E, H, with $H \neq \emptyset$ (even when $\Pi(H) = 0$).

Nevertheless the above definition is given on sets of conditional events endowed with a logical structure, and this can make the model not flexible for the applications. In fact, in any real situation the events of interest, and those in which the field expert or the decision maker has information, give rise usually to an *arbitrary set*. For this reason we study a notion of coherence (see [4]), which allows to see whether in fact a partial assessment is the restriction of a conditional possibility (for $T = \min$) and then to enlarge coherent assessments on any further conditional event. We point out a procedure to check coherence by using a characterization theorem given in [4]. The procedure and the relevant results are naturally generalized in this talk. The relevant theorems characterize coherent T-conditional possibility assessments in terms of a class of possibilities satisfying suitable properties. In [4] an independence notion for conditional possibilities is given and it is shown that it overcomes critical situations; this suggest that an extension of independence for any t-norm can be carried out, analogously to what has been done for strictly increasing t-norms in [13].

In this work, we give a contribution to the discussion about conditioning under a different perspective by studying the comparative framework underling a conditional model. In many situations the field expert or the decision maker, due to his partial knowledge, is not able or interested to give a numerical evaluation "even if partial". In these situations, we are content with getting (from the decision maker) an ordinal evaluation (i.e. a comparative degree of belief among conditional events) comparing only some uncertain alternatives. In this case, given a numerical model of reference (in our case possibilistic framework) it necessary to determine the conditions characterizing ordinal relations $\preceq$, which are representable by a function $\varphi(\cdot|\cdot)$ (e.g. possibility, necessity measures) belonging to the numerical reference model, i.e. for every $E|H, F|K \in \mathcal{A} \times (\mathcal{A} \setminus \{\emptyset\})$

$$\begin{aligned} E|H \preceq F|K &\Rightarrow \varphi(E|H) \leq \varphi(F|K) \\ E|H \prec F|K &\Rightarrow \varphi(E|H) < \varphi(F|K) \end{aligned}.$$

In [5] relations representable by a conditional possibility have been characterized and here we deeply study this class of models. It is interesting in decision theory also since in the unconditional case conditions on acts (in the style of Savage) - assuring a possibilistic representation - have been given in [12] and an optimistic attitude of the decision maker has been carried out. This allows to analyze different notions of conditional possibilities from a qualitative point of view.

In this talk we will emphasize the different conditioning operators in this qualitative framework showing different characterizations and examples.

Moreover, we characterize ordinal relations locally representable, admitting a specific conditional measure (possibility or necessity) $\varphi(\cdot|\cdot)$ such that for every $E, F, H \in \mathcal{A}$, with $H \supseteq E \vee F$ we have:

$$(*) \qquad \begin{aligned} E \preceq F &\Rightarrow \varphi(E|H) \leq \varphi(F|H) \\ E \prec F &\Rightarrow \varphi(E|E \vee F) < \varphi(F|E \vee F) \end{aligned}.$$

This approach gives an estimation of the "goodness" and "effectiveness" of the model, by putting in evidence the rules necessarily accepted by the user. This is particularly useful in the case that there are many numerical models. In fact, it puts clearly in evidence the lacks of some conditioning operators or some unexpected similarities among the conditional measures.

This analysis puts under a new light the conditioning problem for possibility theory: for conditional possibility (with $T = \min$) it comes out that a relation is locally representable if and only if it is representable by a strict positive unconditional measure, as shown by the following result.

Before we recall the following characteristic axiom (PO) (introduced in [10]) for possibility:

$$(PO) \qquad \text{for any } A,B,H \in \mathcal{A},\ A \preceq B \Rightarrow (A \vee H) \preceq (B \vee H).$$

Theorem 1. *Let $\preceq$ be an ordinal relation on the algebra $\mathcal{A}$. Then the following statements are equivalent:*

(a) $\preceq$ is a monotone weak order satisfying axiom (PO) and such that $\emptyset \prec A$, for any $A \in (\mathcal{A} \setminus \emptyset)$;
(b) there exists a conditional possibility $\Pi(\cdot|\cdot) : \mathcal{A} \times (\mathcal{A} \setminus \emptyset) \to [0,1]$ locally representing $\preceq$;
(c) there exists a T-conditional possibility $\Pi(\cdot|\cdot) : \mathcal{A} \times (\mathcal{A} \setminus \emptyset) \to [0,1]$, with T a strictly monotone triangular norm, locally representing $\preceq$;
(d) there exists a strictly positive (unconditional) possibility $\Pi(\cdot) : \mathcal{A} \to [0,1]$ representing $\preceq$.

Thus, the main conditioning operators are not distinguishable (it means they collapse in the same class) in the qualitative framework under local representability.

The above characterization gives rise to a fundamental difference between this measure and probability: ordinal relations representable by a strict positive probability are also locally representable by a conditional probability, while the vice versa is not true.

Some interesting aspects are obtained also by looking to conditional necessities: in this case we obtain not transitive relations. This is not surprising since it happens also in other frameworks.

References

1. B. Bouchon-Meunier, G. Coletti, C. Marsala (2002) Independence and possibilistic conditioning. *Annals of Mathematics and Artificial Intelligence* 35, 107–124.
2. G.Coletti, R.Scozzafava (2001) From conditional events to conditional measures: a new axiomatic approach. *Annals of Mathematics and Artificial Intelligence*, 32, pp. 373-392.
3. G. Coletti, R. Scozzafava (2002) *Probabilistic logic in a coherent setting*. Trends in logic n.15. Dordrecht/Boston/London: Kluwer.
4. G.Coletti, B. Vantaggi (2006) Possibility theory: conditional independence. *Fuzzy Sets and Systems*, 157, pp. 1491-1513.
5. G.Coletti, B. Vantaggi (2007) Comparative models ruled by possibility and necessity: a conditional world. *International Journal of Approximate Reasoning*, to appear.
6. R.T. Cox (1946) Probability, frequency and reasonable expectation. *American Journal of Physics*, 14 (1), pp. 1-13.
7. G. de Cooman (1997) Possibility theory II: Conditional Possibility. *International Journal of General Systems*, 25, pp. 325–351.
8. B. de Finetti (1931) Sul significato soggettivo della probabilità. *Fundamenta Matematicae*, 17, pp. 293–329.
9. D. Dubois, L. Fariñas del Cerro, A. Herzig, H. Prade (1999) Qualitative relevance and independence: a roadmap, Proc. 15 th Int. Joint Conf. on A.I., Nagoya, 62-67. Extended version: A roadmap of qualitative independence. In: Fuzzy Sets, Logics and Reasoning about Knowledge (Dubois, D., Prade, H., Klement, E.P., eds.), Kluwer Academic Publ.,325–350.
10. D. Dubois (1986) Belief structure, possibility theory and decomposable confidence measures on finite sets. *Comput. Artificial Intelligence* 5, pp. 403-416.
11. D. Dubois, L. Fariñas del Cerro, A. Herzig, H. Prade (1994) An ordinal view of independence with application to plausible reasoning, Proc. of the 10th Conf. on Uncertainty in Artificial Intelligence (R. Lopez de Mantaras, D. Poole, eds.), Seattle, WA, July 29-31, pp. 195-203.
12. D. Dubois, H. Prade, R. Sabbadin (2001) Decision-theoretic foundations of qualitative possibility theory. *European Journal of Operational Research*, 128, pp. 459–478.
13. L. Ferracuti, B. Vantaggi (2006) Independence and conditional possibilities for strictly monotone triangular norms. *International Journal of Intelligent Systems*, 21(3), pp. 299-323.
14. L.A. Zadeh (1978) Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1, pp. 3–28.

Statistical Inference About the Means of Fuzzy Random Variables: Applications to the Analysis of Fuzzy- and Real-Valued Data

Ana Colubi

Dpto. de Estadística e I.O. y D.M.
Universidad de Oviedo 33007 Oviedo, Spain
colubi@uniovi.es

1 Introduction

Statistical data are frequently associated with an underlying imprecision due, for instance, to inexactitude in the measuring process, vagueness of the involved concepts or a certain degree of ignorance about the real values. In many cases, such an imprecision can be modelled by means of fuzzy sets in a more efficient way than considering only a single value or category. Thus, these kinds of data are jointly affected by two sources of uncertainty: fuzziness (due to imprecision, vagueness, partial ignorance) and randomness (due to sampling or measurement errors of stochastic nature). Fuzzy Random Variables (FRVs) in Puri & Ralescu's sense [17] has been introduced to model these situations, that is, random mechanisms generating imprecisely-valued data (see, for instance, [4] for a discussion about the integration of Fuzzy Sets and Statistics).

From a statistical point of view, the fuzzy expected value (see [17]) plays an important role as central summary measure. The point estimation of this measure has been one of the first statistical analysis concerning FRVs (see, for instance, [12], [13]). Later, the initial hypothesis testing procedures began to be studied, although they imposed some theoretical/practical constraints (see, for instance, [10], [14], [15] or [6]). The aim of these procedures were to test whether the expected value of a FRV is a given fuzzy set (one-sample test), or whether the expected value of two or more FRVs are equal (two-sample and J-sample tests). In order to solve these problems, the null hypothesis is expressed in terms of suitable metrics (see [1], [10] and [11]). More recently, the constraints have been removed and some operative and powerful bootstrap hypothesis testing were proposed (see, for instance, [8] and [9]).

On the other hand, in [7] (see also [3]) it has been shown that in handling real-valued random variables and by using suitable fuzzifications, the corresponding fuzzy expected value captures not only the "central tendency summary" but the whole information on the distribution of the original variables. In this sense, the fuzzy expected value means a kind of $[0,1]$-valued *characteristic function* and the above-mentioned hypothesis tests concerning the expected value of the fuzzy random variables can be used to develop inferences about distributions of real-valued random variables.

The final aim of this paper is to review the statistical inferences about the fuzzy expected value, from both a theoretical and an empirical point of view, and to show how to apply them in connection with the distributions of real-valued random variables through the characterizing fuzzy representations. In Section 2, we present the main concepts concerning FRV's. In Section 3 we illustrate the hypothesis testing procedures to be analyzed. Finally, in Section 4, we present briefly the characterizing fuzzy representation of real-valued random variables and some of the inferences about the distributions.

2 Preliminaries

Let $\mathcal{K}_c(\mathbb{R}^p)$ be the class of the nonempty compact convex subsets of $\mathbb{R}^p$ endowed with the Minkowski sum and the product by a scalar, that is, $A+B=\{a+b\,|\,a\in A\, b\in B\}$ and $\lambda A=\{\lambda a\,|\,a\in A\}$ for all $A,B\in\mathcal{K}_c(\mathbb{R}^p)$ and $\lambda\in\mathbb{R}$. We will consider the *class of fuzzy sets*

$$\mathcal{F}_c(\mathbb{R}^p)=\big\{U:\mathbb{R}^p\to[0,1]\;\big|\;U_\alpha\in\mathcal{K}_c(\mathbb{R}^p)\text{ for all }\alpha\in[0,1]\big\}$$

where U_α is the α-level of U (i.e. $U_\alpha=\{x\in\mathbb{R}^p\,|\,U(x)\geq\alpha\}$) for all $\alpha\in(0,1]$, and U_0 is the closure of the support of U. The space $\mathcal{F}_c(\mathbb{R}^p)$ can be endowed with the sum and the product by a scalar based on Zadeh's extension principle [18], which satisfies that $(U+V)_\alpha=U_\alpha+V_\alpha$ and $(\lambda U)_\alpha=\lambda U_\alpha$ for all $U,V\in\mathcal{F}_c(\mathbb{R}^p)$, $\lambda\in\mathbb{R}$ and $\alpha\in[0,1]$.

The support function of a fuzzy set $U\in\mathcal{F}_c(\mathbb{R}^p)$ is $s_U(u,\alpha)=\sup_{w\in U_\alpha}\langle u,w\rangle$ for any $u\in\mathbb{S}^{p-1}$ and $\alpha\in[0,1]$, where $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$ and $\langle\cdot,\cdot\rangle$ denotes the inner product. The support function allows to embed $\mathcal{F}_c(\mathbb{R}^p)$ onto a cone of the class of the Lebesgue integrable functions $\mathcal{L}(\mathbb{S}^{p-1})$ by means of the mapping $s:\mathcal{F}_c(\mathbb{R}^p)\to\mathcal{L}(\mathbb{S}^{p-1}\times[0,1])$ where $s(U)=s_U$ (see [5]).

We will consider the *generalized metric* by Körner and Näther [11], which is defined so that

$$[D_K(U,V)]^2=\int_{(\mathbb{S}^{p-1})^2\times[0,1]^2}\big(s_U(u,\alpha)-s_V(u,\alpha)\big)\big(s_U(v,\beta)-s_V(v,\beta)\big)\,dK(u,\alpha,v,\beta),$$

for all $U,V\in\mathcal{F}_c(\mathbb{R}^p)$, where K is a positive definite and symmetric kernel. Thus, D_K coincides with the generic L_2 distance on the Banach space $\mathcal{L}(\mathbb{S}^{p-1}\times[0,1])$.

Let $(\Omega,\mathcal{A},P)$ be a probability space. A *Fuzzy Random Variable* in Puri & Ralescu's sense [16] is a mapping $\mathcal{X}:\Omega\to\mathcal{F}_c(\mathbb{R}^p)$ so that the α-level mappings $\mathcal{X}_\alpha:\Omega\to\mathcal{K}_c(\mathbb{R}^p)$, defined so that $\mathcal{X}_\alpha(\omega)=\big(\mathcal{X}(\omega)\big)_\alpha$ for all $\omega\in\Omega$, are random sets (that is, Borel-measurable mappings with the Borel σ-field generated by the topology associated with the well-known Hausdorff metric d_H on $\mathcal{K}(\mathbb{R}^p)$). Alternatively, an FRV is an $\mathcal{F}_c(\mathbb{R}^p)$-valued random element (i.e. a Borel-measurable mapping) when the D_K-metric is considered on $\mathcal{F}_c(\mathbb{R}^p)$ (see [2], [11] and [5]).

If $\mathcal{X}:\Omega\to\mathcal{F}_c(\mathbb{R}^p)$ is a fuzzy random variable such that $d_H\big(\{0\},\mathcal{X}_0\big)\in L^1(\Omega,\mathcal{A},P)$, then the *expected value (or mean)* of $\mathcal{X}$ is the unique $E(\mathcal{X})\in\mathcal{F}_c(\mathbb{R}^p)$ such that $\big(E(\mathcal{X})\big)_\alpha=$ Aumann's integral of the random set $\mathcal{X}_\alpha$ for all $\alpha\in[0,1]$, that is,

$$\big(E(\mathcal{X})\big)_\alpha=\big\{E(X|P)\;\big|\;X:\Omega\to\mathbb{R}^p,X\in L^1(\Omega,\mathcal{A},P),X\in\mathcal{X}_\alpha\;a.s.\;[P]\big\}.$$

2.1 Inferences on the mean values of FRVs

Let $\mathcal{X}:\Omega\to\mathcal{F}_c(\mathbb{R}^p)$ be a FRV such that $d_H\big(\{0\},\mathcal{X}_0\big)\in L^1(\Omega,\mathcal{A},P)$ and let $\mathcal{X}_1,\ldots,\mathcal{X}_n$ be FRVs which are identically distributed as $\mathcal{X}$. Then (see [13]),

Theorem 1. *The* ***sample fuzzy mean value*** $\overline{\mathcal{X}}_n=\dfrac{1}{n}(\mathcal{X}_1+\ldots+\mathcal{X}_n)$ *is an unbiased and consistent estimator of the fuzzy parameter* $E(\mathcal{X})$*; that is, the (fuzzy) mean of the fuzzy-valued estimator* $\overline{\mathcal{X}}_n$ *over the space of all random samples equals* $E(\mathcal{X})$ *and* $\overline{\mathcal{X}}_n$ *converges almost-surely to* $E(\mathcal{X})$

In addition, if $\mathcal{X}_1^*,\ldots,\mathcal{X}_n^*$ is a bootstrap sample obtained from $\mathcal{X}_1,\ldots,\mathcal{X}_n$. Then (see [7]), we have that

Theorem 2. *In testing the null hypothesis* $H_0 : E(\mathcal{X}|P) = U \in \mathcal{F}_c(\mathbb{R}^p)$ *at the nominal significance level* $\alpha \in [0,1]$, H_0 *should be rejected whenever*

$$\frac{\left[D_K\left(\overline{\mathcal{X}}_n, U\right)\right]^2}{\widehat{S}^2_{K_n}} > z_\alpha,$$

where z_α *is the* $100(1-\alpha)$ *fractile of the distribution of the bootstrap statistic* $T_n = \left[D_K\left(\overline{\mathcal{X}}^*_n, \overline{\mathcal{X}}_n\right)\right]^2 \Big/ \widehat{S}^{*2}_{K_n}$ *and with*

$$\overline{\mathcal{X}}_n = \frac{1}{n}\left(\mathcal{X}_1 + \ldots + \mathcal{X}_n\right), \; \widehat{S}^2_{K_n} = \sum_{i=1}^n \left[D_K\left(\mathcal{X}_i, \overline{\mathcal{X}}_n\right)\right]^2/(n-1),$$
$$\overline{\mathcal{X}}^*_n = \frac{1}{n}\left(\mathcal{X}^*_1 + \ldots + \mathcal{X}^*_n\right), \; \widehat{S}^{*2}_{K_n} = \sum_{i=1}^n \left[D_K(\mathcal{X}^*_i, \overline{\mathcal{X}}^*_n)\right]^2/(n-1).$$

Analogously, testing hypothesis can also be developed for the two (independent or paired)-sample and multi-sample cases (see [6] and [8]).

3 Characterizing fuzzy representations of random variables

A *fuzzy representation of a random variable* transforms crisp data (the original random variable values) into fuzzy sets (the associated FRV values). In [7] a family of fuzzy representations that characterize the distribution of the original real-valued random variable is proposed. Concretely, if we define the generalized mapping $\gamma^C : \mathbb{R} \to \mathcal{F}_c(\mathbb{R})$ which transforms each value $x \in \mathbb{R}$ into the fuzzy number whose α-level sets are $\left(\gamma^C(x)\right)_\alpha = \left[f_L(x) - (1-\alpha)^{1/h_L(x)},\; f_R(x) + (1-\alpha)^{1/h_R(x)}\right]$ for all $\alpha \in [0,1]$, where $f_L : \mathbb{R} \to \mathbb{R}$, $f_R : \mathbb{R} \to \mathbb{R}$, $f_L(x) \leq f_R(x)$ for all $x \in \mathbb{R}$, and $h_L : \mathbb{R} \to (0,+\infty)$, $h_R : \mathbb{R} \to (0,+\infty)$ are continuous and bijective, we have the following result:

Theorem 3. *If* $X : \Omega \to \mathbb{R}$ *and* $Y : \Omega \to \mathbb{R}$ *are two random variables and* $f_L(X), f_R(X) \in L^1(\Omega, \mathcal{A}, P)$, *the two following conditions:*

(C.1) $\widetilde{E}\left(\gamma^C \circ X \,\middle|\, P\right) = \widetilde{E}\left(\gamma^C \circ Y \,\middle|\, P\right)$.
(C.2) *X and Y are identically distributed,*

are equivalent.

As a consequence of Theorem 3, we have that

- the fuzzy sample mean value of the fuzzified random variable becomes an estimator of the distribution;
- the one-sample test about the mean can be considered as a goodness-of-fit test for the original random variable;
- the J-sample test becomes a test for the equality of J distributions.

References

1. C. Bertoluzza, N. Corral, and A. Salas. On a new class of distances between fuzzy numbers. *Mathware Soft Comput.*, 2:71–84, 1995.
2. A. Colubi, J. S. Domínguez-Menchero, M. López-Díaz, and D. A. Ralescu. A $d_e[0,1]$-representation of random upper semicontinuous functions. *Proc. Amer. Math. Soc.*, 130:3237–3242, 2002.

3. A. Colubi and G. González-Rodríguez. Triangular fuzzification of random variables and power of distribution tests: empirical discussion. (accepted, in press), 2006.
4. R. Coppi, M.A. Gil, and H.A.L. Kiers. The fuzzy approach to statistical analysis. *Comput. Statist. Data Anal.*, 51:1–14, 2006.
5. P. Diamond and P. Kloeden. *Metric Spaces of Fuzzy Sets: Theory and Applications*. World Scientific, Singapore, 1994.
6. M.A. Gil, M. Montenegro, G. González-Rodríguez, A. Colubi, and M.R. Casals. Bootstrap approach to the multi-sample test of means with imprecise data. *Comput. Statist. Data Anal.*, 51:148–162, 2006.
7. G. González-Rodríguez, A. Colubi, and M.A. Gil. A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. *Comput. Statist. Data Anal.*, 51:163–176, 2006.
8. G. González-Rodríguez, A. Colubi, M.A. Gil, and P. D'Urso. An asymptotic two dependent samples test of equality of means of fuzzy random variables. In *Proc. COMPSTAT'2006*, pages 689–695, 2006.
9. G. González-Rodríguez, M. Montenegro, A. Colubi, and M.A. Gil. Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems*, 157:2608–2613, 2006.
10. R. Körner. An asymptotic α-test for the expectation of random fuzzy variables. *J. Statist. Plann. Inference*, 83:331–346, 2000.
11. R. Körner and W. Näther. On the variance of random fuzzy variables. In C. Bertoluzza, M.A. Gil, and D.A. Ralescu, editors, *Statistical Modeling, Analysis and Management of Fuzzy Data*, pages 22–39. Physica-Verlag, Heidelberg, 2002.
12. M. A. Lubiano and M. A. Gil. Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Statist. Papers*, 40:277–295, 1999.
13. M. A. Lubiano, M. A. Gil, and M. López-Díaz. On the rao-blackwell theorem for fuzzy random variables. *Kybernetika*, 35:167–175, 1999.
14. M. Montenegro, M. R. Casals, M. A. Lubiano, and M. A. Gil. Two-sample hypothesis tests of means of a fuzzy random variable. *Inform. Sci.*, 133:89–100, 2001.
15. M. Montenegro, A. Colubi, M. R. Casals, and M. A. Gil. Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika*, 59:31–49, 2004.
16. M. L. Puri and D. A. Ralescu. The concept of normality for fuzzy random variables. *Ann. Probab.*, 11:1373–1379, 1985.
17. M. L. Puri and D. A. Ralescu. Fuzzy random variables. *J. Math. Anal. Appl.*, 114:409–422, 1986.
18. L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, I. *Inform. Sci.*, 8:199–249, 1975.

Orthogonal Grid Constructions of Copulas

Bernard De Baets[1] and Hans De Meyer[2]

[1] Department of Applied Mathematics, Biometrics and Process Control,
Ghent University, Coupure links 653, B-9000 Gent, Belgium
bernard.debaets@ugent.be
[2] Department of Applied Mathematics and Computer Science,
Ghent University, Krijgslaan 281 S9, B-9000 Gent, Belgium
A-4040 Linz, Austria
hans.demeyer@ugent.be

We establish a general framework for constructing copulas that can be regarded as a patchwork-like assembly of arbitrary copulas, with non-overlapping rectangles as patches. We derive a family of construction methods that require the choice of a single background copula. When this background copula is the greatest copula $T_{\mathbf{M}}$, we retrieve the well-known ordinal sum construction, while in case of the smallest copula $T_{\mathbf{L}}$, we obtain a construction method that is dual to the ordinal sum construction. Also non-singular background copulas lead to suitable construction methods.

Connecting Two Theories of Imprecise Probability

Gert de Cooman

SYSTeMS Research Group
Ghent University
Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium
gert.decooman@UGent.be

1 Introduction

In recent years, we have witnessed the growth of a number of theories of uncertainty, where imprecise (lower and upper) probabilities, or probability intervals, rather than precise (or point-valued) probabilities, have a central part. Here we consider two of them, Peter Walley's behavioural theory [8], and Glenn Shafer and Vladimir Vovk's game-theoretic account of probability [7]. These seem to have a completely different interpretation, and they certainly stem from quite different schools of thought: Walley follows the tradition of Frank Ramsey [6], Bruno de Finetti [3] and Peter Williams [10] in trying to establish a rational model for a subject's beliefs in terms of her behaviour. Shafer and Vovk follow an approach that has many other influences as well, and is strongly coloured by ideas about gambling systems and martingales. They use Cournot's bridge to interpret lower and upper probabilities (see [7, Chapter 2] for a nice historical overview), whereas on Walley's approach, lower and upper probabilities are defined in terms of a subject's betting rates.

What we set out to do here, is show that in many practical situations, both approaches are very strongly connected. This means that results, valid in one theory, can automatically be converted and reinterpreted in terms of the other.

2 Shafer and Vovk's game-theoretic approach to probability

In their game-theoretic approach to probability [7], Shafer and Vovk consider a game with two players, World and Skeptic, who play according to a certain *protocol*. They obtain the most interesting results for a special type of protocol, called a *coherent probability protocol*. This section is devoted to explaining what this means.

G1. The first player, World, can make a number of moves, where the possible next moves may depend on the previous moves he has made, but do not in any way depend on the previous moves made by Skeptic.

This means that we can represent his game-play by a (decision) tree. We restrict ourselves here to the discussion of *bounded protocols*, where World can only make a finite and bounded number of moves, whatever happens. But we do not exclude the possibility that at some point in the tree, World has the choice between an infinite number of next moves.

Let us establish some terminology related to World's decision tree. A *path* in the tree represents a possible sequence of moves for World from the beginning to the end of the game. We denote the set of all possible paths ω by Ω, the *sample space* of the game. A *situation* t is some connected segment of a path that is *initial*, i.e., starts at the root of the tree. It identifies the decisions or moves World has made up to a certain point. We denote the set of all situations by $\Omega^{\Diamond}$. It includes the set Ω of *final* situations, or paths. All other situations are called *non-final*; among them is the *initial* situation

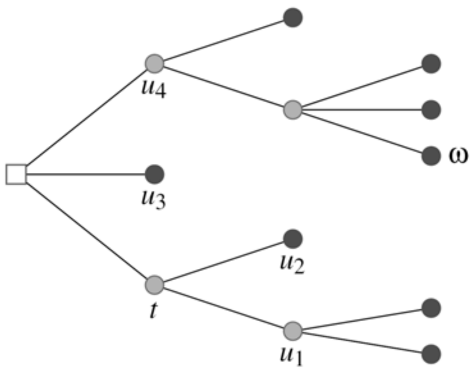

Fig. 1. A simple decision tree for World, displaying the initial situation □, other non-final situations (such as t) as grey circles, and paths, or final situations, (such as ω) as black circles. Also depicted is a cut of □, consisting of the situations u_1, ..., u_4.

□, which represents the empty initial segment. See Figure 1 for a graphical example explaining these notions.

World's *move space* in a non-final situation t is the set $\mathbf{W}_t$ of those moves $\mathbf{w}$ that World can make in t: $\mathbf{W}_t = \{\mathbf{w}\colon t\mathbf{w} \in \Omega^\diamond\}$.

If for two situations s and t, s is an (initial) segment of t, then we say that s *precedes* t or that t *follows* s, and write $s \sqsubseteq t$. If ω is a path and $t \sqsubseteq \omega$ then we say that the path ω *goes through* situation t. We write $s \sqsubset t$ if $s \sqsubseteq t$ and $s \neq t$.

A function on $\Omega^\diamond$ is called a *process*, and a partial process whose domain includes all situations that follow a situation t is called a *t-process*. Similarly, a function on Ω is called a *variable*, and a partial variable on Ω whose domain includes all paths that go through a situation t is called a *t-variable*. If we restrict a t-process $\mathcal{F}$ to the final situations that follow t, we obtain a t-variable, which we denote by $\mathcal{F}_\Omega$.

We now turn to the other player, Skeptic. His moves may be influenced by the previous moves that World has made, in the following sense. In each situation t, he has a set $\mathbf{S}_t$ of moves $\mathbf{s}$ available to him, called Skeptic's *move space* in t.

G2. In each non-final situation t, there is a (positive or negative) gain for Skeptic associated with each of the possible moves $\mathbf{s}$ in $\mathbf{S}_t$ that World can make. This gain depends only on the situation t and the next move $\mathbf{w}$ that World will make.

This means that for each non-final situation t there is a *gain function* $\lambda_t\colon \mathbf{S}_t \times \mathbf{W}_t \to \mathbb{R}$, such that $\lambda_t(\mathbf{s},\mathbf{w})$ represents the change in Skeptic's capital in situation t when he makes move $\mathbf{s}$ and World makes move $\mathbf{w}$.

Let us introduce some further notions and terminology related to Skeptic's game-play. A *strategy* $\mathcal{P}$ for Skeptic is a partial process defined on the set $\Omega^\diamond \setminus \Omega$ of non-final situations, such that $\mathcal{P}(t) \in \mathbf{S}_t$ is the move that Skeptic will make in each non-final situation t. With each such strategy $\mathcal{P}$ there corresponds a *capital process* $\mathcal{K}^\mathcal{P}$, whose value in each situation t gives us Skeptic's capital accumulated so far, when he starts out with zero capital and plays according to the strategy $\mathcal{P}$. It is given by the recursion relation

$$\mathcal{K}^\mathcal{P}(t\mathbf{w}) = \mathcal{K}^\mathcal{P}(t) + \lambda_t(\mathcal{P}(t),\mathbf{w}),$$

with initial condition $\mathcal{K}^{\mathcal{P}}(\square) = 0$. Of course, when Skeptic starts out (in $\square$) with capital α and uses strategy $\mathcal{P}$, his corresponding accumulated capital is given by the process $\alpha + \mathcal{K}^{\mathcal{P}}$. In the final situations, his accumulated capital is then given by the variable $\alpha + \mathcal{K}^{\mathcal{P}}_{\Omega}$.

If we start in a non-final situation t, rather than in $\square$, then we can consider t-strategies $\mathcal{P}$ that tell Skeptic how to move starting from t, and the corresponding capital process $\mathcal{K}^{\mathcal{P}}$ is then also a t-process, that tells us how much capital Skeptic has accumulated since starting with zero capital in situation t and using t-strategy $\mathcal{P}$.

Assumptions G1 and G2 determine so-called *gambling protocols*. They are sufficient for us to be able to define upper and lower prices for variables. Consider a non-final situation t and a t-variable f. Then the *upper price* $\overline{\mathbb{E}}_t(f)$ *for* f *in* t is defined as the infimum capital α that Skeptic has to start out with in t in order that there would be some t-strategy $\mathcal{P}$ such that his accumulated capital $\alpha + \mathcal{K}^{\mathcal{P}}$ allows him, at the end of the game, to buy f, whatever moves World makes after t:

$$\overline{\mathbb{E}}_t(f) := \inf\left\{\alpha \colon \text{there is some } t\text{-strategy } \mathcal{P} \text{ such that } \alpha + \mathcal{K}^{\mathcal{P}}_{\Omega} \geq f\right\}, \tag{1}$$

where $\alpha + \mathcal{K}^{\mathcal{P}}_{\Omega} \geq f$ is taken to mean that $\alpha + \mathcal{K}^{\mathcal{P}}(\omega) \geq f(\omega)$ for all final situations ω that go through t. Similarly, for the *lower price* $\underline{\mathbb{E}}_t(f)$ *for* f *in* t:

$$\underline{\mathbb{E}}_t(f) := \sup\left\{\alpha \colon \text{there is some } t\text{-strategy } \mathcal{P} \text{ such that } \alpha - \mathcal{K}^{\mathcal{P}}_{\Omega} \leq f\right\},$$

so $\underline{\mathbb{E}}_t(f) = -\overline{\mathbb{E}}_t(-f)$. If we start from the initial situation $t = \square$, we simply get the *upper and lower prices* for a variable f, which we also denote by $\overline{\mathbb{E}}(f)$ and $\underline{\mathbb{E}}(f)$.

A gambling protocol is called a *probability protocol* when besides S1 and S2, two more requirements are satisfied.

P1. For each non-final situation t, Skeptic's move space $\mathbf{S}_t$ is a convex cone in some linear space: $a_1\mathbf{s}_1 + a_2\mathbf{s}_2 \in \mathbf{S}_t$ for all non-negative real numbers a_1 and a_2 and all $\mathbf{s}_1$ and $\mathbf{s}_2$ in $\mathbf{S}_t$.

P2. For each non-final situation t, Skeptic's gain function λ_t has the following linearity property: $\lambda_t(a_1\mathbf{s}_1 + a_2\mathbf{s}_2, \mathbf{w}) = a_1\lambda_t(\mathbf{s}_1, \mathbf{w}) + a_2\lambda_t(\mathbf{s}_2, \mathbf{w})$ for all non-negative real numbers a_1 and a_2, all $\mathbf{s}_1$ and $\mathbf{s}_2$ in $\mathbf{S}_t$ and all $\mathbf{w}$ in $\mathbf{W}_t$.

Finally, a probability protocol is called *coherent* when moreover

C. For each non-final situation t, and for each $\mathbf{s}$ in $\mathbf{S}_t$ there is some $\mathbf{w}$ in $\mathbf{W}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) \leq 0$.

It is clear what this requirement means: for each non-final situation, World has a strategy for playing from t onwards such that Skeptic cannot (strictly) increase his capital from t onwards, whatever t-strategy he uses.

For such coherent probability protocols, Shafer and Vovk prove a number of interesting properties for the corresponding upper (and lower) prices. We list a number of them here. Call a *cut* U of a non-final situation t any set of situations that (i) follow t, and (ii) such that for all paths ω through t [$t \sqsubseteq \omega$], there is a unique $u \in U$ such that ω goes through u [$u \sqsubseteq \omega$]; see also Figure 1. For any t-variable f, we can associate with such a cut U another special t-variable $\overline{\mathbb{E}}_U$ by $\overline{\mathbb{E}}_U(f)(\omega) = \overline{\mathbb{E}}_u(f)$, for all paths ω through t, where u is the unique situation in U that ω goes through. For any two t-variables f_1 and f_2, $f_1 \leq f_2$ is taken to mean that $f_1(\omega) \leq f_2(\omega)$ for all paths ω that go through t.

Proposition 1 (Properties of prices in a coherent probability protocol [7]). *Consider a coherent probability protocol, let* t *be a non-final situation,* f*,* f_1 *and* f_2 t*-variables, and* U *a cut of* t*. Then*

1. $\inf\{f(\omega) \colon \omega \in \Omega, t \sqsubseteq \omega\} \leq \underline{\mathbb{E}}_t(f) \leq \overline{\mathbb{E}}_t(f) \leq \sup\{f(\omega) \colon \omega \in \Omega, t \sqsubseteq \omega\}$ *[positivity];*

2. $\overline{\mathbb{E}}_t(f_1+f_2) \leq \overline{\mathbb{E}}_t(f_1)+\overline{\mathbb{E}}_t(f_2)$ *[sub-additivity];*
3. $\overline{\mathbb{E}}_t(\lambda f) = \lambda\overline{\mathbb{E}}_t(f)$ *for all real* $\lambda \geq 0$ *[non-negative homogeneity];*
4. $\overline{\mathbb{E}}_t(f+\alpha) = \overline{\mathbb{E}}_t(f)+\alpha$ *for all real* α *[constant additivity];*
5. $\overline{\mathbb{E}}_t(\alpha) = \alpha$ *for all real* α *[normalisation];*
6. $f_1 \leq f_2$ *implies that* $\overline{\mathbb{E}}_t(f_1) \leq \overline{\mathbb{E}}_t(f_2)$ *[monotonicity];*
7. $\overline{\mathbb{E}}_t(f) = \overline{\mathbb{E}}_t(\overline{\mathbb{E}}_U(f))$ *[law of iterated expectation].*

What is more, Shafer and Vovk use specific instances of such coherent probability protocols to prove various limit theorems (such as the law of large numbers, the central limit theorem, the law of the iterated logarithm), from which they can derive the well-known measure-theoretic versions. We shall come back to this in Section 5.

3 Walley's behavioural approach to probability

In his book on the behavioural theory of imprecise probabilities [8], Walley considers many different types of related models. We shall restrict ourselves here to the most general and most powerful one, which also turns out to be the easiest to explain, namely coherent sets of desirable gambles; see also [9].

Consider a non-empty set Ω of possible alternatives ω, only one which actually obtains (or will obtain); we assume that it is possible, at least in principle, to determine which alternative does so. Also consider a subject who is uncertain about which possible alternative actually obtains (or will obtain). A *gamble* on Ω is a real-valued function on Ω, and it is interpreted as an uncertain reward, expressed in units of some predetermined linear utility scale: if ω actually obtains, then the reward is $f(\omega)$, which may be positive or negative. If a subject *accepts* a gamble f, this means that she is willing to engage in the transaction, where (i) first it is determined which ω obtains, and then (ii) she receives the reward $f(\omega)$. We can try and model the subject's beliefs about Ω by considering which gambles she accepts.

Suppose our subject specifies some set $\mathcal{D}$ of gambles she accepts, called a *set of desirable gambles*. Such a set is called *coherent* if it satisfies the following *rationality requirements*:

D1. if $f < 0$ then $f \notin \mathcal{D}$ [avoiding partial loss];
D2. if $f \geq 0$ then $f \in \mathcal{D}$ [accepting partial gain];
D3. if f_1 and f_2 belong to $\mathcal{D}$ then their (point-wise) sum $f_1 + f_2$ also belongs to $\mathcal{D}$ [combination];
D4. if f belongs to $\mathcal{D}$ then its (point-wise) scalar product λf also belongs to $\mathcal{D}$ for all non-negative real numbers λ [scaling].

Here '$f < 0$' means '$f \leq 0$ and not $f = 0$'. Walley has also argued that sets of desirable gambles should satisfy an additional axiom:

D5. $\mathcal{D}$ is $\mathcal{B}$-conglomerable for any partition $\mathcal{B}$ of Ω: if $I_B f \in \mathcal{D}$ for all $B \in \mathcal{B}$, then also $f \in \mathcal{B}$ [full conglomerability].

Full conglomerability is a very strong requirement, and it is not without controversy. If a model $\mathcal{D}$ is $\mathcal{B}$-conglomerable, this means that certain inconsistency problems when conditioning on elements B of $\mathcal{B}$ are avoided; see [8] for more details. Conglomerability of belief models was not required by forerunners of Walley, such as Williams [10], or de Finetti [3]. While I agree with Walley that conglomerability is a desirable property for sets of desirable gambles, I do not believe that *full* conglomerability is always necessary: it seems that we need only require conglomerability with respect to those partitions that we actually intend to condition our model on. This is the path I shall follow in Section 4.

Given a coherent set of desirable gambles, we can define *conditional upper and lower previsions* as follows: for any gamble f and any non-empty subset B of Ω, with indicator I_B,

$$\overline{P}(f|B) := \inf\{\alpha\colon I_B(\alpha - f) \in \mathcal{D}\}$$
$$\underline{P}(f|B) := \sup\{\alpha\colon I_B(f - \alpha) \in \mathcal{D}\}$$

so $\underline{P}(f|B) = -\overline{P}(-f|B)$, and $\overline{P}(f|B)$ is the infimum price α for which the subject will sell the gamble f, i.e., accept the gamble $\alpha - f$, contingent on the occurrence of B. For any event A, we define the conditional lower probability $\underline{P}(A|B) := \underline{P}(I_A|B)$, i.e., the subject's supremum rate for betting on the event A, contingent on the occurrence of B, and similarly for $\overline{P}(A|B) := \overline{P}(I_A|B)$.

If $\mathcal{B}$ is a partition of Ω, then we define $\overline{P}(f|\mathcal{B})$ as the gamble that in any element ω of Ω assumes the value $\overline{P}(f|B)$, where B is the unique element of $\mathcal{B}$ such that $\omega \in B$.

The following properties of conditional upper and lower previsions associated with a coherent set of desirable gambles were (essentially) proven by Walley.

Proposition 2 (Properties of conditional upper and lower previsions [8]). *Consider a coherent set of desirable gambles, let B be any non-empty subset of Ω, f, f_1 and f_2 gambles on Ω,. Then*[1]

1. $\inf\{f(\omega)\colon \omega \in B\} \leq \underline{P}(f|B) \leq \overline{P}(f|B) \leq \sup\{f(\omega)\colon \omega \in B\}$ *[positivity];*
2. $\overline{P}(f_1 + f_2|B) \leq \overline{P}(f_1|B) + \overline{P}(f_2|B)$ *[sub-additivity];*
3. $\overline{P}(\lambda f|B) = \lambda\overline{P}(f|B)$ *for all real* $\lambda \geq 0$ *[non-negative homogeneity];*
4. $\overline{P}(f + \alpha|B) = \overline{P}(f|B) + \alpha$ *for all real* α *[constant additivity];*
5. $\overline{P}(\alpha|B) = \alpha$ *for all real* α *[normalisation];*
6. $f_1 \leq f_2$ *implies that* $\overline{P}(f_1|B) \leq \overline{P}(f_2|B)$ *[monotonicity];*
7. *if $\mathcal{B}$ is a partition of Ω that refines the partition $\{B, B^c\}$ and $\mathcal{D}$ is $\mathcal{B}$-conglomerable, then* $\overline{P}(f|B) \leq \overline{P}(\overline{P}(f|\mathcal{B})|B)$ *[conglomerative property].*

The analogy between Propositions 1 and 2 is too striking to be coincidental. The fact that there is an equality in Proposition 1.7, where we have only an inequality in Propositions 2.7, seems to indicate moreover that Shafer and Vovk's approach leads to a less general type of model.[2] We now set out to identify the exact correspondence between the two models.

4 Connecting the two approaches

In order to lay bare the connections between the game-theoretic and the behavioural approach, we enter Shafer and Vovk's world, and consider another player, called Subject, who has certain *piece-wise* beliefs about what moves World will make.

More specifically, for each non-final situation $t \in \Omega^\diamond \setminus \Omega$, she has beliefs about which move $\mathbf{w}$ World will choose next from the set $\mathbf{W}_t$ of moves available to him in t. We suppose she represents

[1] Here, as in Proposition 1, we assume that whatever we write down is well-defined, meaning that for instance no sums of $-\infty$ and $+\infty$ appear, and that the function $\overline{P}(\cdot|\mathcal{B})$ is real-valued, and nowhere infinite. Shafer and Vovk do not seem to mention this.

[2] This also shows that the claim on p. 186 in [7] to the effect that "[de Finetti, Williams and Walley] also considered the relation between unconditional and conditional prices, but they were not working in an dynamic framework and so did not formulate [Shafer and Vovk's equivalent of our Proposition 1.7]", at least needs some qualification.

those beliefs in the form of a coherent[3] set $\mathcal{D}_t$ of desirable gambles on $\mathbf{W}_t$. These beliefs are conditional, in the sense that they represent Subject's beliefs about what World will do *immediately after he gets to situation t*. We call any specification of such coherent $\mathcal{D}_t$, $t \in \Omega^\lozenge \setminus \Omega$, a *coherent conditional assessment* for Subject.

We can now ask ourselves what the behavioural implications of these conditional assessments are. For instance, what do they tell us about whether or not Subject should accept certain gambles on Ω, the set of possible paths for World? In other words, how can these beliefs about which next move World will make in each non-final situation t be combined rationally into beliefs about World's complete sequence of moves?

In order to investigate this, we use Walley's very general and powerful method of *natural extension*, which is just *conservative coherent reasoning*. We shall construct, using the pieces of information $\mathcal{D}_t$, a set of desirable gambles on Ω that is (i) coherent, and (ii) as small as possible, meaning that no more gambles should be accepted than is actually required by coherence.

First, we collect the pieces. Consider any non-final situation $t \in \Omega^\lozenge \setminus \Omega$ and any gamble h_t in $\mathcal{D}_t$. Just as for variables, we can define a t-gamble as a partial gamble whose domain contains all paths ω that go through t. Then with each h_t we can associate a t-gamble, also denoted by h_t, and defined by

$$h_t(\omega) := h_t(\omega(t)),$$

for all $t \sqsubseteq \omega$, where we denote by $\omega(t)$ the unique element of $\mathbf{W}_t$ such that $t\omega(t) \sqsubseteq \omega$. If we consider the set $\uparrow t := \{\omega \in \Omega \colon t \sqsubseteq \omega\}$ of all paths that go through t, then $I_{\uparrow t} h_t$ represents the gamble on Ω that is called off unless World ends up in situation t, and which, when it is not called off, depends only on World's move immediately after t, and gives the same value $h_t(\mathbf{w})$ to all paths ω that go through $t\mathbf{w}$. The fact that Subject accepts h_t on $\mathbf{W}_t$ contingent on World's getting to t, translates immediately to the fact that Subject accepts the gamble $I_{\uparrow t} h_t$ on Ω. We thus end up with a set of gambles on Ω

$$\mathcal{D} := \bigcup_{t \in \Omega^\lozenge \setminus \Omega} \{I_{\uparrow t} h_t \colon h_t \in \mathcal{D}_t\}$$

that Subject accepts. The only thing left to do now, is to find the smallest coherent set $\mathcal{E}_\mathcal{D}$ of desirable gambles that includes $\mathcal{D}$ (if there is such a coherent set). Here we take coherence to refer to conditions D1–D4, together with D5', a variation on D5 which refers to conglomerability with respect to those partitions that we actually intend to condition on, as discussed in Section 3.

These partitions are what we call cut partitions. Consider any *non-final* cut $U \subseteq \Omega^\lozenge \setminus \Omega$ of the initial situation $\square$. Then the set of events $\mathcal{B}_U := \{\uparrow u \colon u \in U\}$ is a partition of Ω, called the *U-partition*. D5' requires that our set of desirable gambles should be *cut conglomerable*, i.e., conglomerable with respect to every cut partition $\mathcal{B}_U$.

Because we require cut conglomerability, it follows that $\mathcal{E}_\mathcal{D}$ will contain the sums of gambles $\sum_{u \in U} I_{\uparrow u} h_u$ for all non-final cuts U of $\square$ and all choices of $h_u \in \mathcal{D}_u$, $u \in U$. Because $\mathcal{E}_\mathcal{D}$ should be a convex cone [by D3 and D4], any sum of such sums $\sum_{u \in U} I_{\uparrow u} h_u$ over a finite number of non-final cuts U should also belong to $\mathcal{E}_\mathcal{D}$. But, since in the case of bounded protocols we are discussing here, World can only make a bounded and finite number of moves, $\Omega^\lozenge \setminus \Omega$ is a finite union of such non-final cuts, and therefore the sums $\sum_{u \in \Omega^\lozenge \setminus \Omega} I_{\uparrow u} h_u$ should belong to $\mathcal{E}_\mathcal{D}$ for all choices $h_u \in \mathcal{D}_u$, $u \in \Omega^\lozenge \setminus \Omega$.

Call therefore, for any initial situation t, a *t-selection* any partial process $\mathcal{S}$ defined on the non-final situations $s \sqsupseteq t$ such that $\mathcal{S}(s) \in \mathcal{D}_s$. With such a t-selection, we can associate a t-process, called a

[3] Since we do not envisage conditioning this model on subsets of $\mathbf{W}_t$, we impose no extra conglomerability requirements here, only the coherence conditions D1–D4.

gamble process $\mathcal{G}^{\mathcal{S}}$, with value

$$\mathcal{G}^{\mathcal{S}}(s) = \sum_{t \sqsubseteq u, u \sqsubset s} I_{\uparrow u}\mathcal{S}(u)(s(u))$$

in all situations s that follow t. Alternatively, $\mathcal{G}^{\mathcal{S}}$ is given by the recursion relation $\mathcal{G}^{\mathcal{S}}(s\mathbf{w}) = \mathcal{G}^{\mathcal{S}}(s) + \mathcal{S}(s)(\mathbf{w})$ for all non-final $s \sqsupseteq t$, with initial value $\mathcal{G}^{\mathcal{S}}(t) = 0$. In particular, this leads to the t-gamble $\mathcal{G}^{\mathcal{S}}_{\Omega}$ defined on all final situations ω that follow t, by letting

$$\mathcal{G}^{\mathcal{S}}_{\Omega} = \sum_{t \sqsubseteq u, u \in \Omega^{\diamond} \setminus \Omega} I_{\uparrow u}\mathcal{S}(u).$$

Then we have just shown that the gambles $\mathcal{G}^{\mathcal{S}}_{\Omega}$ should belong to $\mathcal{E}_{\mathcal{D}}$ for all non-final situations t and all t-selections $\mathcal{S}$. As before for strategy and capital processes, we call a $\square$-selection $\mathcal{S}$ simply a *selection*, and a $\square$-gamble process simple a *gamble process*. It is now but a small step to prove the following result.

Proposition 3. *The smallest set of gambles that satisfies D1–D4 and D5' and includes $\mathcal{D}$, or in other words, the natural extension of $\mathcal{D}$, is given by*

$$\mathcal{E}_{\mathcal{D}} := \left\{ g \colon \textit{there is some selection } \mathcal{S} \textit{ such that } g \geq \mathcal{G}^{\mathcal{S}}_{\Omega} \right\}.$$

Moreover, for any non-final situation t and any t-gamble g, we have that $I_{\uparrow t}g \in \mathcal{E}_{\mathcal{D}}$ if and only if $g \geq \mathcal{G}^{\mathcal{S}}_{\Omega}$ for some t-selection $\mathcal{S}$, where as before, $g \geq \mathcal{G}^{\mathcal{S}}_{\Omega}$ is taken to mean that $g(\omega) \geq \mathcal{G}^{\mathcal{S}}_{\Omega}(\omega)$ for all final situations ω that follow t.

We now use the coherent set of desirable gambles $\mathcal{E}_{\mathcal{D}}$ to define upper (and lower) previsions conditional on the cut partitions $\mathcal{B}_U$ as indicated in Section 3. We then get, using Proposition 3, that for any cut U of $\square$ and any situation u in U:

$$\begin{aligned} \overline{P}(f|\uparrow u) &:= \inf \left\{ \alpha \colon I_{\uparrow u}(\alpha - f) \in \mathcal{E}_{\mathcal{D}} \right\} \\ &= \inf \left\{ \alpha \colon \text{there is some } u\text{-selection } \mathcal{S} \text{ such that } \alpha - \mathcal{G}^{\mathcal{S}}_{\Omega} \geq f \right\}. \end{aligned} \tag{2}$$

There seems to be a close correspondence between the expressions [such as (1)] for upper prices $\overline{\mathbb{E}}_t(f)$ associated with coherent probability protocols and those [such as (2)] for the conditional upper previsions $\overline{P}(f|\uparrow t)$ based on a coherent conditional assessments. This correspondence is made explicit in the following theorem. Say that a given coherent probability protocol and given coherent conditional assessment *match* whenever they lead to identical corresponding upper prices $\overline{\mathbb{E}}_t$ and conditional upper previsions $\overline{P}(\cdot|\uparrow t)$ for all non-final $t \in \Omega^{\diamond} \setminus \Omega$.

Theorem 1 (Matching Theorem). *For every coherent probability protocol there is a coherent conditional assessment such that both match, and conversely, for every coherent conditional assessment there is a coherent probability protocol such that both match.*

The proof of this result is quite technical, but the underlying ideas should be clear. If we have a coherent probability protocol with move spaces $\mathbf{S}_t$ and gain functions λ_t for Skeptic, define the conditional assessment for Subject to be (essentially) $\mathcal{D}_t := \{-\lambda(\mathbf{s}, \cdot) \colon \mathbf{s} \in \mathbf{S}_t\}$. If, conversely, we have a coherent conditional assessment for Subject consisting of the sets $\mathcal{D}_t$, define the move spaces for Skeptic by $\mathbf{S}_t := \mathcal{D}_t$, and his gain functions by $\lambda_t(h, \cdot) := -h$ for all h in $\mathcal{D}_t$.

5 Interpretation

The Matching Theorem has a very interesting interpretation. In Shafer and Vovk's approach, World is sometimes decomposed into two players, Reality and Forecaster. It is Reality whose moves are characterised by the above-mentioned decision tree, and it is Forecaster who determines in each non-final situation t what Skeptic's move space $\mathbf{S}_t$ and gain function λ_t is. We now go beyond Shafer and Vovk's model, by adding something to it.

Suppose that Forecaster has certain beliefs, in each non-final situation t, about what move Reality will make next, and suppose she models those beliefs by specifying a coherent set $\mathcal{D}_t$ of desirable gambles on $\mathbf{W}_t$. In other words, we identify Forecaster with Subject.

When Forecaster specifies such a set, she is making certain behavioural commitments. In fact, she is committing herself to accepting any gamble in $\mathcal{D}_t$, and to accepting any combination of such gambles according to the combination axioms D3, D4 and D5'. This implies that we can derive conditional upper previsions $\overline{P}(\cdot|\uparrow t)$, with the following interpretation: in situation t, $\overline{P}(f|\uparrow)$ is the infimum price for which Forecaster can be made to sell the t-gamble f for on the basis of the commitments she has made.

What Skeptic can now do, is take Forecaster up on her commitments. This means that in each situation t, he can select a gamble (or equivalently, any non-negative linear combination of gambles) h_t in $\mathcal{D}_t$ and offer it to Forecaster. If Reality's next move in situation t is $\mathbf{w} \in \mathbf{W}_t$, this means that Skeptic can increase his capital by (the positive or negative amount) $-h_t(\mathbf{w})$, by exploiting Forecaster's commitments. In other words, his move space $\mathbf{s}_t$ can then be identified with the convex set of gambles $\mathcal{D}_t$ and his gain function λ_t is then given by $\lambda_t(h_t,\cdot) = -h_t$. But then Theorem 1 tells us that this leads to a coherent probability protocol, and that the corresponding upper prices $\overline{\mathbb{E}}_t$ for Skeptic coincide with Forecaster's conditional upper previsions $\overline{P}(\cdot|\uparrow t)$.

This is of particular relevance to the laws of large numbers that Shafer and Vovk derive in their game-theoretic framework, because such laws now can be given a behavioural interpretation in terms of Forecaster (or any Subject's) (conditional) lower and upper previsions. To give an example, let us consider the following game.

FINITE-HORIZON BOUNDED FIXED LOWER FORECASTING GAME

Parameters: N, $B > 0$, $\varepsilon > 0$, $\alpha > 0$
Players: Reality, Forecaster, Skeptic
Protocol: Forecaster announces $m \in [-B, B]$
  $\mathcal{K}_0 := \alpha$
  FOR $n = 1, \ldots, N$:
    Skeptic announces $\lambda_n \geq 0$
    Reality announces $x_n \in [-B, B]$
    $\mathcal{K}_n = \mathcal{K}_{n-1} + \lambda_n(x_n - m)$.
Winner: Skeptic wins if $\mathcal{K}_n$ is never negative and either $\mathcal{K}_N \geq 1$ or $\frac{1}{N}\sum_{n=1}^{N} x_n < m - \varepsilon$. Otherwise Reality wins.

Then Enrique Miranda and I have proven elsewhere [1], amongst other things, that Skeptic has a strategy that guarantees that he wins the game if he starts out with capital $\alpha \geq \exp(-\frac{N\varepsilon^2}{16B^2})$. But this means that for the event

$$\Delta_{N,\varepsilon} := \left\{ (x_1, \ldots, x_N) \colon \frac{1}{N} \sum_{n=1}^{N} x_n < m - \varepsilon \right\}$$

we have that $\overline{\mathbb{E}}(\Delta_{N,\varepsilon}) \leq \exp(-\frac{N\varepsilon^2}{16B^2})$. We are now able to import this result into the behavioural theory of imprecise probabilities, using Theorem 1. Consider a number of bounded random variables $X_1, \ldots,$ X_N, where $X_k \in [-B, B]$, whose values will be revealed successively. Assume that some Subject models her beliefs about the values that these variables assume by specifying, on beforehand, a common lower prevision m for each of them, meaning that she accepts to buy each X_k for any price that is at least m.[4] Then coherence requires her to bet on the event that the sample mean $\frac{1}{N}\sum_{n=1}^{N} X_k$ will be at least $m - \varepsilon$ at rates that are higher than $1 - \exp(-\frac{N\varepsilon^2}{16B^2})$, so these rates go to one as N increases, for any $\varepsilon > 0$. This is a weak law of large numbers for bounded random variables.

6 Additional Remarks

We have proven the correspondence between the two approaches only for decision trees with a bounded horizon. For games with infinite horizon, the correspondence becomes less immediate, because Shafer and Vovk implicitly make use of coherence axioms that are stronger than D1–D4 and D5', leading to upper prices that are dominated by the corresponding conditional upper previsions. Exact matching would be restored of course, provided we can argue that these additional requirements are rational for any subject to comply with. This could be an interesting topic for further research.

References

1. G. de Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 2006. Submitted for publication.
2. B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.
3. B. de Finetti. *Theory of Probability*. John Wiley & Sons, Chichester, 1974–1975. English translation of [2], two volumes.
4. P. Gärdenfors and N.-E. Sahlin. *Decision, Probability, and Utility*. Cambridge University Press, Cambridge, 1988.
5. H. E. Kyburg Jr. and H. E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.
6. F. P. Ramsey. Truth and probability (1926). In R. B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, chapter VII, pages 156–198. Kegan, Paul, Trench, Trubner & Co., London, 1931. Reprinted in [5] and [4].
7. G. Shafer and V. Vovk. *Probability and Finance: It's Only a Game*. Wiley, New York, 2001.
8. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
9. P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.
10. P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.

[4] Observe that this implies that Subject is not learning about the value of X_k from previous observations of $X_1, \ldots, X_{k-1}$, because she uses the same m, independently of what has happened before. This means that there is some assessment of independence, which we have called [1] forward irrelevance, which is much weaker than the usual independence assessment found in more common weak laws of large numbers.

On the Relationships Between Random Sets, Possibility Distributions, P-Boxes and Clouds

Sebastien Destercke[1,2], Didier Dubois[2], and Eric Chojnacki[1]

[1] Institute of Radiological Protection and Nuclear Safety (IRSN)
Cadarache, France
{sebastien.destercke,eric.chojnacki}@irsn.fr
[2] Toulouse Institute of Research in Computer Science (IRIT)
31062 Toulouse, France
{Destercke,Dubois}@irit.fr

There are many practical representations of probability families that make them easier to handle in applications. Among them are random sets, possibility distributions, Ferson's p-boxes [4] and Neumaier's clouds [6]. Both for theoretical and practical considerations, it is very useful to know whether one representation can be translated into or approximated by other ones. We first briefly recall formalisms and existing results, before exhibiting relationships between all these representations. In this note, which is a summary of an extended forthcoming paper, we restrict ourselves to representations on a finite set $X = \{x_1, x_2, \dots, x_n\}$ of n elements.

1 Formalisms

Possibility distribution A possibility distribution is a mapping $\pi : X \to [0,1]$ representing incomplete information about an ill-known parameter v. two dual measures (respectively the possibility and necessity measures) can be defined : $\Pi(A) = \sup_{x \in A} \pi(x)$ and $N(A) = 1 - \Pi(A^c)$. To any normal possibility distribution (π such that $\pi(x) = 1$ for some $x \in X$) can be associated a probability family $\mathcal{P}_\pi$ s.t. $\mathcal{P}_\pi = \{P, \forall A \subseteq X \text{ measurable}, N(A) \leq P(A) \leq \Pi(A)\}$.

Random Set A random set is defined here as a probability distribution on the power set of X, namely $m : 2^X \to [0,1]$. $m(A)$ is the probability that all is known about v is that $v \in A$. Two dual measures (respectively the plausibility and belief measures) can be defined : $Pl(A) = \sum_{E, E \cap A \neq \emptyset} m(E)$ and $Bel(A) = 1 - Pl(A^c) = \sum_{E, E \subseteq A} m(E)$. To any random set m can be associated a probability family $\mathcal{P}_m$ s.t. $\mathcal{P}_m = \{P | \forall A \subseteq X \text{ measurable}, Bel(A) \leq P(A) \leq Pl(A)\}$.

Generalized p-box A p-box is usually defined on the real line by a pair of cumulative distributions $[\underline{F}, \overline{F}]$, defining the probability family $\mathcal{P}_{[\underline{F},\overline{F}]} = \{P | \underline{F}(x) \leq F(x) \leq \overline{F}(x) \quad \forall x \in \Re\}$. The notion of cumulative distribution on the real line is based on a natural ordering of numbers. In order to generalize this notion to arbitrary finite sets, we need to define a weak order relation $\leq_R$ on this space. Given $\leq_R$, an R-downset is of the form $\{x_i : x_i \leq_R x\}$, and denoted $(x]_R$. A generalized R-cumulative distribution is defined as the function $F_R : X \to [0,1]$ s.t. $F_R(x) = \Pr((x]_R)$, where Pr is a probability measure on X. We can now define a generalized p-box as a pair $[\underline{F}_R(x), \overline{F}_R(x)]$ of generalized cumulative distributions defining a probability family $\mathcal{P}_{[\underline{F}_R(x),\overline{F}_R(x)]} = \{P | \forall x, \underline{F}_R(x) \leq F_R(x) \leq \overline{F}_R(x)\}$. Generalized P-boxes can also be represented by a set of constraints

$$\alpha_i \leq P(A_i) \leq \beta_i \qquad i = 1, \dots, n \tag{1}$$

where $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n \leq 1$, $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_n \leq 1$ and $A_i = (x_i]_R, \forall x_i \in X$ with $x_i \leq_R x_j$ iff $i < j$ (sets A_i form a sequence of nested confidence sets $\emptyset \subset A_1 \subset A_2 \subset \ldots \subset A_n \subset X$).

Cloud Formally, a cloud is described by an Interval-Valued Fuzzy Set (IVF) s.t. $(0,1) \subseteq \cup_{x\in X} F(x) \subseteq [0,1]$, where $F(x)$ is an interval $[\delta(x), \pi(x)]$. A cloud is called *thin* when the two membership functions coincide ($\delta = \pi$). It is called *fuzzy* when the lower membership function δ is 0 everywhere. Let α_i be a sequence of α-cuts s.t. $1 = \alpha_0 > \alpha_1 > \alpha_2 > \ldots > \alpha_n > \alpha_{n+1} = 0$ with A_i, B_i the corresponding α-cut of fuzzy sets π and δ ($A_i = \{x_i, \pi(x_i) > \alpha_{i+1}\}$ and $B_i = \{x_i, \delta(x_i) \geq \alpha_{i+1}\}$). Then, a random variable x is in a cloud if it satisfies the constraints

$$P(B_i) \leq 1 - \alpha_i \leq P(A_i) \text{ and } B_i \subseteq A_i \quad i = 1, \ldots, n. \tag{2}$$

2 Generalized p-boxes

First, let us notice that a generalized upper cumulative distribution $\overline{F}_R$ can be seen as a possibility distribution π_R dominating a probability distribution Pr, since it is a maxitive measure s.t. $\max_{x\in A} \overline{F}_R(x) \geq \Pr(A), \forall A \subseteq X$. In [2], we have shown the following results

Proposition 1. *A family $\mathcal{P}_{[\underline{F}_R(x), \overline{F}_R(x)]}$ described by a generalized P-box can be encoded by a pair of possibility distributions π_1, π_2 s.t. $\mathcal{P}_{[\underline{F}_R(x), \overline{F}_R(x)]} = \mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$ with $\pi_1(x) = \overline{F}_R(x)$ and $\pi_2(x) = 1 - \underline{F}_R(x)$*

Proposition 2. *A family $\mathcal{P}_{[\underline{F}_R(x), \overline{F}_R(x)]}$ described by a generalized P-box can be encoded by a random set m s.t. $\mathcal{P}_{[\underline{F}_R(x), \overline{F}_R(x)]} = \mathcal{P}_m$.*

If X is the real line, this last proposition reduces to results already shown in [5].

3 Cloud

In [3], the following relationship linking clouds to possibility distributions is shown

Proposition 3. *A probability family $\mathcal{P}_{\delta,\pi}$ described by the cloud (δ, π) is equivalent to the family $\mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$ described by the two possibility distributions π and $1 - \delta$.*

This result already suggests that clouds and generalized p-boxes are somewhat related. To lay bare this relationship, it is useful to introduce the following special case of clouds:

Definition 1. *A cloud is said to be* comonotonic *if distributions π and δ are comonotonic. If it is not the case, a cloud is called non-comonotonic.*

Remark 1. Thin and *fuzzy* clouds are special cases of comonotonic clouds.

Figure 1 illustrates comonotonic and non-comonotonic clouds. The two following propositions show why it is useful to make this distinction.

Proposition 4. *The probability family $\mathcal{P}_{\delta,\pi}$ induced by a comonotonic cloud is equivalent to a generalized p-box and can thus be encoded through a random set.*

Proof (sketch). Since comonotonicity imply that sets A_i, B_i $i = 1, \ldots, n$ form a complete sequence of nested sets, one can always retrieve the structure of a generalized p-box from a comonotonic cloud by mapping constraints of the form of equation (2) into constraints of the form of equation (1).

Proposition 5. *The lower probability of the probability family $\mathcal{P}_{\delta,\pi}$ induced by a non-comonotonic cloud is not even a 2-monotone capacity (i.e. $\exists A, B \subset X$ s.t. $\underline{P}(A \cap B) + \underline{P}(A \cup B) \leq \underline{P}(A) + \underline{P}(B)$)*

Fig. 1. Illustration of clouds.

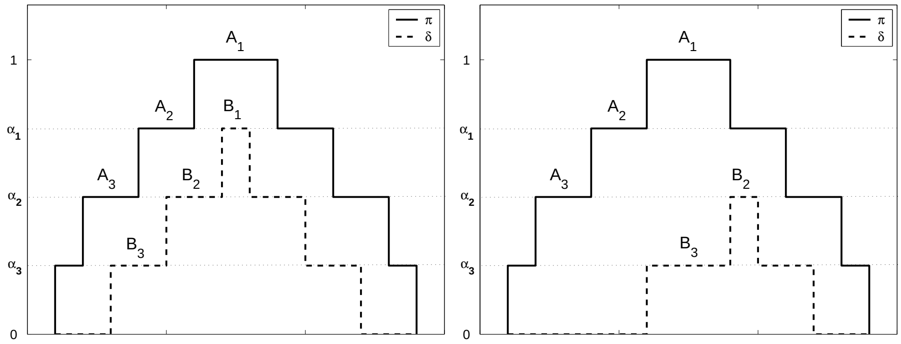

Fig 1.A Comonotonic cloud　　Fig 1.B Non-comonotonic cloud

Proof (sketch). For each non-comonotonic cloud, there exist two sets B_i, A_j with $i > j$ and s.t. $B_i \cap A_j \neq \emptyset$, $B_i \not\subseteq A_j$ and $A_j \not\subseteq B_i$. Using a result from Chateauneuf [1] and the fact that $\mathcal{P}_{\delta,\pi}$ is the intersection of two families corresponding to belief functions, we can show that the following inequality holds

$$\underline{P}(A_j \cap B_i) + \underline{P}(A_j \cup B_i) < \underline{P}(A_j) + \underline{P}(B_i)$$

and this concludes the proof.

Remark 2. δ must not be trivially reduced to a single set B_n s.t. $B_n \cap A_{n-1} = \emptyset$, otherwise the cloud can still be encoded by a random set (and is thus a capacity of order ∞), even if it is no longer equivalent to a generalized p-box.

To our knowledge, non-comonotonic clouds are the only simple models (in the finite case, we need at most $2|X|$ values to fully specify a cloud) of imprecise probabilities that induce capacities that are not 2-monotone.

Let us also notice that if proposition 4 holds in the continuous case, we have a nice way to characterize probability families induced by comonotonic clouds. Namely, a continuous belief function [7] with uniform mass density, whose focal elements would be disjoint sets of the form $[x(\alpha), u(\alpha)] \cup [v(\alpha), y(\alpha)]$ where $\{x : \pi(x) \geq \alpha\} = [x(\alpha), y(\alpha)]$ and $\{x : \delta(x) \geq \alpha\} = [u(\alpha), v(\alpha)]$. In particular, for *thin* clouds, focal sets would be doubletons of the form $\{x(\alpha), y(\alpha)\}$.

Computing upper and lower probability bounds $\underline{P}(A), \overline{P}(A)$ of non-comonotonic clouds appear not to be so easy a task. Thus, one may wish to work with inner or outer approximations of the family $\mathcal{P}_{\delta,\pi}$. The two following propositions provide such bounds, which are easy to compute.

Proposition 6. *If $\mathcal{P}_{\delta,\pi}$ is the probability family described by the cloud (δ, π) on a referential X, then, the following bounds provide an outer approximation :*

$$\max(N_\pi(A), N_\delta(A)) \leq P(A) \leq \min(\Pi_\pi(A), \Pi_\delta(A)) \ \forall A \subset X \tag{3}$$

Remark 3. These bounds are the ones considered by Neumaier in [6], and the fact that they are outer approximations explain why they are poorly related to random sets or to Walley's natural extensions.

But clouds can be approximated by random sets:

Proposition 7. *Given sets* $\{B_i, A_i, i = 1, \ldots, n\}$ *and the corresponding confidence values* α_i*, associated to the distributions* (δ, π) *of a cloud, the belief and plausibility measures of the random set s.t.* $m(A_i \setminus B_{i-1}) = \alpha_{i-1} - \alpha_i$ *are inner approximations of* $\mathcal{P}_{\delta,\pi}$.

Remark 4. If the cloud is comonotonic, this random set is the one corresponding to the family $\mathcal{P}_{\delta,\pi}$

Our results show that clouds generalize p-boxes and possibility distributions as representations of imprecise probabilities, but are generally not a special case of random set. Even if they look more complex to deal with than p-boxes and possibility distributions, clouds are more expressive and remain relatively simple representations. Moreover, results presented here may allow for easier computations in various cases. We thus think that using clouds can be potentially interesting in various applications, but that more work is needed to fully assess this potential.

4 Open questions and problems

There remain many open questions and problems related to clouds, some of them being already emphasized by Neumaier. Among them are :

- Testing the mathematical and the computational tractability of clouds
- Testing clouds as descriptive models of uncertainty
- Extending existing results to more general frameworks (unbounded variables, lower/upper previsions)
- Studying under which operations the cloud representation is preserved (joint distributions, fusion, extension, . . .)

Acknowledgements

Authors wish to thank Matthias Troffaes for illuminating discussions on properties of clouds. This abstract has been supported by a grant from the Institute of Radiological Protection and Nuclear Safety (IRSN). Scientific responsibility rests with the authors.

References

1. Alain Chateauneuf. Combination of compatible belief functions and relation of specificity. In *Advances in the Dempster-Shafer theory of evidence*, pages 97–114. John Wiley & Sons, Inc, New York, NY, USA, 1994.
2. S. Destercke and D. Dubois. A unified view of some representations of imprecise probabilities. In J. Lawry, E. Miranda, A. Bugarin, and S. Li, editors, *Int. Conf. on Soft Methods in Probability and Statistics (SMPS)*, Advances in Soft Computing, pages 249–257, Bristol, 2006. Springer.
3. D. Dubois and H. Prade. Interval-valued fuzzy sets, possibility theory and imprecise probability. In *Proceedings of International Conference in Fuzzy Logic and Technology (EUSFLAT'05)*, Barcelona, September 2005.
4. S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, and K. Sentz. Construction probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.
5. E. Kriegler and H. Held. Utilizing belief functions for the estimation of future climate change. *I. J. of Approximate Reasoning*, 39:185–209, 2005.
6. A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.
7. P. Smets. Belief functions on real numbers. *I. J. of Approximate Reasoning*, 40:181–223, 2005.

A New Family of Symmetric Copulas

Fabrizio Durante

Department of Knowledge-Based Mathematical Systems
Johannes Kepler University, A-4040 Linz, Austria
fabrizio.durante@jku.at

1 Introduction

For many years, statisticians have been fascinated by the following problem: given n univariate d.f.'s $F_1, F_2, \dots, F_n$, $n \geq 2$, find an n–dimensional d.f. H having F_i as its margins, and having useful properties like a simple analytic expression, a statistical interpretation, and a given depedence structure. Many methods and procedures for constructing such joint distributions have been introduced and studied in the literature [5, 7]. Thanks to Sklar's Theorem [11], this problem can be reduced into the construction of new *copulas*, i.e. multivariate d.f.'s whose univariate margins are uniformly distributed on $[0,1]$. For more details, we refer to [6, 10].

Many families of copulas have been recently introduced in literature, but not all of them are "good" families, in the sense that they can be useful in certain statistical applications. In [6, 8], some desirable properties for a parametric family of bivariate copulas $\{C_\alpha\}_\alpha$, where α belongs to an interval of the real line, are listed and are here reproduced:

(a) *interpretability*, which means having a statistical interpretation;
(b) *flexible and wide range of dependence*, which implies that the copula $\Pi(x,y) = xy$ and at least one of the Fréchet–Hoeffding bounds $W(x,y) = \max(x+y-1,0)$ and $M(x,y) = \min(x,y)$ belong to the class;
(c) *closed form*, in the sense that every member of the family is absolutely continuous or has a simple representation;
(d) *extendibility*, in the sense that the family admits a multivariate extension to the n–dimensional case, $n \geq 3$.

This note aims to present a new family of bivariate copulas sharing all the above properties. All the presented results can be found in [2–4].

2 Characterization of the new class

Given $f : [0,1] \to [0,1]$, we consider the function C_f defined, for every $x, y \in [0,1]$, by

$$C_f(x,y) = (\min(x,y)) f(\max(x,y)). \tag{1}$$

Obviously, every C_f is symmetric, viz. $C(x,y) = C(y,x)$ for every x and y in $[0,1]$, and the copulas Π and M can be represented in the form (1): it suffices to take, respectively, $f(t) = t$ and $f(t) = 1$ for all $t \in [0,1]$. Our aim is to study under which conditions on f, C_f is a copula.

Theorem 1. *Let $f : [0,1] \to [0,1]$ be a continuous function. Let C_f be the function defined by* (1)*. Then C_f is a copula if, and only if, the following statements hold:*

(i) $f(1) = 1$*;*

(ii) *f is increasing;*
(iii) *the function $t \mapsto f(t)/t$ is decreasing on* $]0,1]$.

A function f that satisfies the assumptions of Theorem 1 is called *generator* of a copula of type (1). The class of all generators is convex and, because of condition (iii), it has minimal element $id_{[0,1]}$ and maximal element the constant function equal to 1.

In the sequel we give some sub–families of copulas $\{C_\alpha\}$ of type (1) generated by a one–parameter family $\{f_\alpha\}$ of generators.

Example 1 (Fréchet copulas). Given the generator $f_\alpha(t) := \alpha t + (1-\alpha)$ ($\alpha \in [0,1]$), we obtain $C_\alpha = \alpha\Pi + (1-\alpha)M$, which is a convex sum of Π and M and, therefore, is a member of the Fréchet family of copulas (see[6, family B11]). Notice that $C_0 = M$ and $C_1 = \Pi$.

Example 2 (Cuadras–Augé copulas). Given the generator $f_\alpha(t) := t^\alpha$ ($\alpha \in [0,1]$), C_α is defined by

$$C_\alpha(x,y) = \begin{cases} xy^\alpha, & \text{if } x \le y; \\ x^\alpha y, & \text{if } x > y. \end{cases}$$

Then C_α describes the Cuadras–Augé family of copulas [1].

Example 3. Given the generator $f_\alpha(t) := \min(\alpha t, 1)$ ($\alpha \ge 1$), C_α is defined by

$$C_\alpha(x,y) = \begin{cases} \alpha xy, & \text{if } (x,y) \in [0,1/\alpha]^2; \\ \min(x,y), & \text{otherwise;} \end{cases}$$

viz. C_α is the ordinal sum $(\langle 0, 1/\alpha, \Pi\rangle)$.

3 Properties of this new class

For a copula C_f of type (1) the following statistical interpretation holds [9].

Theorem 2. *If C_f is the copula given by* (1) *and $H(x,y) = C_f(F_1(x), F_2(y))$ for univariate d.f.'s F_1 and F_2, then the following statements are equivalent:*

(a) *random variables X and Y with joint d.f. H have a representation of the form*

$$X = \max\{R, W\} \qquad \textit{and} \qquad Y = \max\{S, W\}$$

where R, S and W are independent r.v.'s;

(b) *H has the form $H(x,y) = F_R(x)F_S(y)F_W(\min(x,y))$, where F_R, F_S and F_W are univariate d.f.'s.*

Moreover, the concordance ordering in this family (which is equivalent to the pointwise ordering between real functions) can be expressed by means of the generators.

Proposition 1. *Let C_f and C_g be two copulas of type* (1) *generated, respectively, by f and g. Then $C_f \le C_g$ if, and only if, $f(t) \le g(t)$ for all $t \in [0,1]$.*

Example 4. Consider the family $\{f_\alpha\}_{\alpha \ge 1}$ given by $f_\alpha(t) := 1 - (1-t)^\alpha$. It is easily proved by differentation that every f_α is increasing with $f_\alpha(t)/t$ decreasing on $]0,1]$. Therefore, this family generates a family of copulas C_α, that is positively ordered, viz. $C_{\alpha_1} \le C_{\alpha_2}$ for $\alpha_1 \le \alpha_2$, with $C_1 = \Pi$ and $C_\infty = M$.

In particular, for every copula C_f, $C_f \geq \Pi$ and, hence, every C_f is positively quadrant dependent [10]. Moreover, every copula of type (1), except Π, has a singular component along the main diagonal of the unit square.

Remark 1. In [3], the authors proved that a copula can be expressed in the form (1) if, and only if, it is *semilinear*, viz. the mappings $h: [0,x] \to [0,1]$, $h(t) := C(t,x)$, and $v: [0,x] \to [0,1]$, $v(t) := C(x,t)$, are linear for all $x \in]0,1]$.

4 A multivariate extension of the new class

Now, we shall consider n–dimensional extension ($n \geq 3$) of the family of bivariate copulas given by (1). Specifically, given a continuous and increasing functon $f\colon [0,1] \to [0,1]$, we define the mapping $C_f^n: [0,1]^n \to [0,1]$ given by

$$C_f^n(u_1, u_2, \cdots, u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}), \tag{2}$$

where $u_{[1]}, \cdots, u_{[n]}$ denote the components of $(u_1, u_2, \cdots, u_n) \in [0,1]^n$ rearranged in increasing order, i.e. $u_{[1]} = \min(u_1, u_2, \cdots, u_n)$ and $u_{[n]} = \max(u_1, u_2, \cdots, u_n)$. It is easy to note that this expression reduces to (1) in the bivariate case. The following result characterizes the copulas of type (2).

Theorem 3. *Let $f\colon [0,1] \longrightarrow [0,1]$ be a continuous function and let C_f^n be the function defined by* (2)*. Then C_f^n is an n-copula if, and only if,*

(i) $f(1) = 1$;
(ii) *f is increasing;*
(iii) *the function $t \to f(t)/t$ is decreasing on* $(0,1]$.

Example 5. Let α be in $[0,1]$ and consider the function $f(t) = \alpha t + \overline{\alpha}$, with $\overline{\alpha} := 1 - \alpha$. Then, the n-copula C_f^n, denoted by C_α, is given, for every $u_1, u_2, \ldots, u_n$ in $[0,1]$, by

$$C_\alpha(u_1, u_2, \ldots, u_n) = u_{[1]} \prod_{i=2}^{n} (\alpha u_{[i]} + \overline{\alpha}).$$

In particular, for $n = 2$, we obtain the family in Example 1.

Example 6. Let α be in $[0,1]$ and consider the function $f(t) = t^\alpha$. Then, we have that the n–copula defined by (2) is given by

$$C_\alpha(\mathbf{u}) = (\min(u_1, u_2, \ldots, u_n))^{1-\alpha} \prod_{i=1}^{n} u_i^\alpha,$$

and it can be considered as a generalization of the Cuadras-Augé family of bivariate copulas (Example 2). Notice that every copula C_α is a *multivariate extreme copula*, viz. for every $t > 0$ $C_\alpha(u_1^t, u_2^t, \ldots, u_n^t) = C_\alpha^t(u_1, u_2, \ldots, u_n)$ [6].

Now, we give a statistical interpretation for copulas of type (2). Let $W_1, W_2, \ldots, W_n$, Z be $n+1$ independent random variables such that, for all $i \in \{1, 2, \ldots, n\}$, W_i has d.f. f satisfying parts (i), (ii) and (iii) in Theorem 3, and Z has d.f. $g(t) = t/f(t)$. Note that $g(1) = 1$ and g is increasing since $f(t)/t$ is decreasing. Consider the random variables $U_i = \max(W_i, Z)$, for all $i = 1, 2, \ldots, n$. Then, for every $(u_1, u_2, \ldots, u_n)$, the d.f. of the random vector $(U_1, U_2, \ldots, U_n)$ is given by

$$P(U_1 \leq u_1, \ldots, U_n \leq u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}),$$

and, hence, it is a copula of type (2).

References

1. C.M. Cuadras, J. Augé, A continuous general multivariate distribution and its properties, Comm. Statist. - Theory Methods 10 (1981) 339–353.
2. F. Durante, A new family of symmetric bivariate copulas, C.R. Acad. Sci. Paris, Ser. I, in press.
3. F. Durante, R. Mesiar, A. Kolesárová, C. Sempi, Semilinear copulas, submitted.
4. F. Durante, J.J. Quesada-Molina, M. Úbeda-Flores, A new class of multivariate distributions, submitted.
5. T.P. Hutchinson and C.D. Lai, *Continuous bivariate distributions. Emphasising applications,* Rumsby Scientific Publishing, Adelaide, 1990.
6. H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, New York, 1997.
7. N.L. Johnson and S. Kotz, *Distributions in statistics: continuous multivariate distributions*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York-London-Sydney, 1972.
8. G. Kilmedorf and A.R. Sampson, Uniform representations of bivariate distributions, *Comm. Statist. Theory Meth.* **4**, 293–301 (1975).
9. A.W. Marshall, Copulas, marginals and joint distributions, in: *Distribution Functions with Fixed Marginals and Related Topics* (L. Rüschendorf, B. Schweizer and M.D. Taylor, Eds.), Institute of Mathematical Statistics (Lecture Notes – Monograph Series Volume 28), Hayward CA, 1996, pp. 213–222.
10. R.B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.
11. A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

Conjunctors and Their Residual Implicators

Fabrizio Durante[1], Erich Peter Klement[1], Radko Mesiar[2,3], and Carlo Sempi[4]

[1] Department of Knowledge-Based Mathematical Systems
Johannes Kepler University, A-4040 Linz, Austria
{fabrizio.durante|ep.klement}@jku.at
[2] Department of Mathematics and Descriptive Geometry
Faculty of Civil Engineering, Slovak Technical University
81 368 Bratislava, Slovakia
[3] Institute for Research and Applications of Fuzzy Modeling
University of Ostrava, 701 03 Ostrava, Czech Republic
mesiar@math.sk
[4] Dipartimento di Matematica "E. De Giorgi"
University of Lecce, I-73100 Lecce, Italy
carlo.sempi@unile.it

1 Introduction

A fuzzy logic is usually considered as a many-valued propositional logic in which the class of truth values is modelled by the unit interval $[0,1]$, and which forms an extension of the classical boolean logic (see [5, 8]).

In these logics, a key role is played by the so called *conjunctor*, i.e. an operation on $[0,1]$ that is used to extend a boolean conjunction from $\{0,1\}$ into $[0,1]$. Usually, conjuctors are assumed to be *triangular norms*, i.e. monotone binary operation on $[0,1]$ that are associative and commutative (see [9]), which have been largely and fruitfully used in many engineering applications.

However, such functions have strong properties that, in some cases, could represent a limitation to their use. As underlined in [4], for example, "if one works with binary conjunctions and there is no need to extend them for three or more arguments, as happens e.g. in the inference pattern called generalized modus ponens, associativity of the conjunction is an unnecessarily restrictive condition". On the other hand, for example, recent research has showed that non commutative logics can be relevant also in some applications to computer science and fuzzy logic programming (see [6] and the references therein).

Motivated by these considerations, in this note we present various classes of conjunctors and, then, we study residual implicators associated to them.

2 Definitions

A mapping $C : [0,1]^2 \to [0,1]$ is a *conjunctor*, also called *semicopula* ([3]) or *t-seminorm* ([15]), if it satisfies following properties:

(C1) $C(x,1) = C(1,x)$ for every x in $[0,1]$;
(C2) C is increasing in each component.

In particular, a conjunctor C is an extension of the classical Boolean conjunction; in fact, $C(x,y)$ takes values in $[0,1]$, and

$$C(0,0) = C(0,1) = C(1,0) = 0, \qquad C(1,1) = 1.$$

In the sequel, we are mainly interested in *left–continuous* conjunctors, viz. conjunctors that are left–continuous in the both components, and, thus, jointly left–continuous ([9, Proposition 1.19]).

A conjunctor C is a *triangular norm* (briefly, *t–norm*) if it is an associative and commutative operation on $[0,1]$ ([1, 9]).

A conjunctor C is a *quasi–copula* if it is 1–*Lipschitz*, viz. it satisfies

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

for every x_1,x_2,y_1,y_2 in $[0,1]$ ([7, 14]).

A conjunctor C is a *copula* if it is 2–*increasing*, viz.

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

for every x_1,x_2,y_1,y_2 in $[0,1]$, $x_1 \leq x_2$ and $y_1 \leq y_2$ ([13, 14]).

Every copula is a quasi–copula, but the converse is in general not true. However, if we consider associative functions, we have the following result due to Moynihan ([9, Theorem 9.10]).

Proposition 1. *For a t–norm T, T is quasi–copula if, and only if, T is a copula.*

We denote by $\mathcal{L}$, $\mathcal{T}$, $\mathcal{Q}$ and $\mathcal{C}$, respectively, the class of left–continuous conjunctors, t–norms, quasi–copulas and copulas. We have that $\mathcal{T}$, $\mathcal{Q}$ and $\mathcal{C}$ are proper subsets of $\mathcal{L}$. Moreover $\mathcal{C}$ is a proper subset of $\mathcal{Q}$ and, as shown, $\mathcal{T}\cap\mathcal{Q}=\mathcal{T}\cap\mathcal{C}$.

3 Characterizations of *R*–implicators

In the construction of fuzzy logic taking truth values in the interval $[0,1]$, if the interpretation of the conjunction is given by a conjunctor C, then the interpretation $R:[0,1]^2\to[0,1]$ of the implication connective (if no additional logical connectives are given) is usually derived from C by means of the *adjointness condition*

$$C(x,z) \leq y \Longleftrightarrow z \leq R(x,y) \tag{1}$$

for each x, y and z in $[0,1]$. In order to guarantee that this adjointness condition determines the operation R uniquely, one has to suppose that C is left–continuous (see [5, 8]). In this case, $R=R_C$ is given by

$$R_C(x,y)=\sup\{z\in[0,1] \mid C(x,z)\leq y\}, \tag{2}$$

which is called *residual implicator* of C (or C–residuum, shortly).

The characterization of the residual implicator of any left–continuous conjunctor is given here.

Theorem 1. *Let C be a left–continuous conjunctor. Then the C–residuum R_C satisfies the following properties:*

(R1) *$R_C(x,y)=1$ if, and only if, $x\leq y$;*
(R2) *$R_C(1,y)=y$ for all y in $[0,1]$;*
(R3) *R_C is decreasing in the first component;*
(R4) *R_C is increasing in the second component;*
(R5) *R_C is left–continuous in its first component;*
(R6) *R_C is right–continuous in its second component.*

We denote by $\mathcal{R}$ the class of all $R : [0,1]^2 \to [0,1]$ satisfying (R1)–(R6), which are called simply residual implicators. We call *residuation* the mapping $\Psi : \mathcal{L} \to \mathcal{R}$ given, for each C in $\mathcal{L}$, by $\Psi(C) = R_C$, where R_C is defined by (2). The mapping $\Psi : \mathcal{L} \to \mathcal{R}$ is bijective and its inverse Ψ^{-1}, called *deresiduation*, is given, for each R in $\mathcal{R}$, by the mapping $C_R : [0,1]^2 \to [0,1]$ defined by

$$C_R(x,y) = \inf\{R(x,z) \geq y \mid z \in [0,1]\}.$$

For each $C \in \mathcal{L}$, $\Psi(C)$ is the C–residuum, and, for each $R \in \mathcal{R}$ we call $\Psi^{-1}(R)$ the *deresiduum* of R (briefly, *R–deresiduum*).

Proposition 2. *If $R : [0,1]^2 \to [0,1]$ satisfies* (R1)–(R6)*, then the R–deresiduum C_R is a left–continuous conjunctor.*

Now, given $\mathcal{A} \subseteq \mathcal{L}$, we characterize the set $\Psi(\mathcal{A})$.

Theorem 2. *Let C be a left–continuous commutative conjunctor. Then the C–residuum R_C satisfies* (R1)–(R6) *and*

(R7) *for every x,y,z in $[0,1]$, $R_C(x,z) \geq y \Longleftrightarrow R_C(y,z) \geq x$.*

Conversely, if R is in $\mathcal{R}$ and satisfies (R7)*, then the R–deresiduum C_R is a left–continuous commutative conjunctor.*

Theorem 3. *Let C be a left–continuous associative conjunctor. Then the C–residuum R_C satisfies* (R1)–(R6) *and*

(R8) *for every x,y,z,u in $[0,1]$, $R_C(y,R_C(x,z)) \geq u$ if, and only if, there exists v in $[0,1]$ such that $R_C(x,v) \geq y$ and $R_C(v,u) \geq z$.*

Conversely, if R is in $\mathcal{R}$ and satisfies (R8)*, then the R–deresiduum C_R is a left–continuous associative conjunctor.*

In particular, we derive the following characterization of left–continuous t–norms.

Theorem 4. *Let T be a left–continuous t–norms. Then the T–residuum R_T satisfies* (R1)–(R8)*. Conversely, if R is in $\mathcal{R}$ and satisfies* (R7) *and* (R8)*, then the R–deresiduum T_R is a left–continuous t–norm.*

Notice that, as a Corollary, we obtain the characterization of the residual implicator of a left–continuous t–norm T given in [12], where (R7) and (R8) are replaced by the condition

$$R_T(x,R_T(y,z)) = R_T(y,R_T(x,z))$$

for each x,y,z in $[0,1]$ (see also [10, 11]). A characterization of the residuum of a continuous t–norm is given by [2].

Theorem 5. *Let Q be a quasi–copula. Then the Q–residuum R_Q satisfies* (R1)–(R6) *and the two following properties:*

(R9) *for every $\varepsilon > 0$, $R_Q(x+\varepsilon,y) \geq R_Q(x,y-\varepsilon)$;*

(R10) *for every $\varepsilon > 0$, $R_Q(x,y) \geq R_Q(x,y-\varepsilon)+\varepsilon$.*

Conversely, if R is an R–implicator satisfying (R9)–(R10)*, then the R–deresiduum Q_R is a quasi–copula.*

Now, by using Proposition 1, we obtain the following result:

Theorem 6. *If C is an associative copula, then the C–residuum R_C satisfies* (R1)–(R10)*. Conversely, if R satisfies* (R1)–(R10)*, then the R–deresiduum C_R is an associative copula.*

Notice that the characterization of the residual implicator of a copula C is still an open question.

4 Induced construction methods for residual implicators

In this section, we investigate whether there exists a close relationship between some construction methods in the class of left–continuous conjunctors and some construction methods in the class of residual implicators.

Proposition 3. *Let $(]a_\alpha, e_\alpha[)_{\alpha\in A}$ be a family of non–empty, pairwise disjoint open subintervals of $[0,1]$ and let $(C_\alpha)_{\alpha\in A}$ be a family of left–continuous conjunctors and . Let C be the ordinal sum $C = (\langle a_\alpha, e_\alpha, C_\alpha\rangle)_{\alpha\in A}$. Then the C–residuum is given by*

$$R_C(x,y) = \begin{cases} a_\alpha + (e_\alpha - a_\alpha)\, R_{C_\alpha}\left(\dfrac{x-a_\alpha}{e_\alpha - a_\alpha}, \dfrac{y-a_\alpha}{e_\alpha - a_\alpha}\right), & a_\alpha < y < x \le e_\alpha; \\ R_{T_{\mathbf{M}}}, & \text{otherwise.} \end{cases} \tag{3}$$

Conversely, if an R–implicator R can be expressed in the form (3)*, then the R–deresiduum C_R is an ordinal sum of the type $C = (\langle a_\alpha, e_\alpha, C_\alpha\rangle)_{\alpha\in A}$.*

Proposition 4. *Let C be a left continuous conjunctor and let φ be an increasing bijection of* $[0,1]$*. Let C_φ be the φ-transform of C given by $C_\varphi(x,y) = \varphi^{-1}(C(\varphi(x),\varphi(y)))$. Then the residual implicator of C_φ is given by*

$$R_{C_\varphi}(x,y) = (R_C)_\varphi(x,y) = \varphi^{-1}(R_C(\varphi(x),\varphi(y))). \tag{4}$$

Conversely, given, an R–implicator R and an increasing bijection φ of $[0,1]$*, consider the φ–transform of R given by $R_\varphi(x,y) = \varphi^{-1}(R(\varphi(x),\varphi(y)))$. Then the R_φ–deresiduum is the φ–transform of the R–deresiduum.*

Proposition 5. *Let C_1 and C_2 be left continuous conjunctors and let C be the pointwise maximum of C_1 and C_2. Then the C–residuum R_C is the pointwise minimum of R_{C_1} and R_{C_2}. Conversely, given two R–implicator R_1 and R_2, let R be the pointwise minimum of R_1 and R_2. Then the R–deresiduum is the pointwise maximum of C_{R_1} and C_{R_2}.*

A similar result holds by replacing maximum with minimum.

Unfortunatelly, the above method cannot be generalized to any pointwise composition of two conjunctors, which is not a lattice operation. For example, given C_1 and C_2 in $\mathcal{L}$, we know that, for any λ in $[0,1]$, $C = \lambda C_1 + (1-\lambda C_2)$ is also in $\mathcal{L}$ and $R = \lambda R_{C_1} + (1-\lambda)R_{C_2}$ is a residual implicator. However, the deresiduum of R, in general, differs from C.

However, this fact stimulate another kind of investigation. We can consider some construction methods directly in the class of residual implicators and then, by deresiduation, we construct new conjunctors, as the followign result shows.

Proposition 6. *Let Q_1 and Q_2 be in $\mathcal{Q}$ and let R_{Q_1} and R_{Q_2} be the R–implicators associated to Q_1 and Q_2, respectively. Then, for every $\lambda \in [0,1]$, $R = \lambda R_{Q_1} + (1-\lambda R_{Q_2})$ is also an R–implicator satisfying (R9)–(R10). Moreover, the deresiduum Q_R is a quasi–copula.*

Example 1. Let $R_{T_{\mathbf{P}}}$ be the residual implicator of $T_{\mathbf{P}}$, $T_{\mathbf{P}}(x,y) = xy$, and let R_C be the residual implicator of the copula C given by $C(x,y) = \sqrt{T_{\mathbf{P}}(x,y)T_{\mathbf{M}}(x,y)}$. Then consider the residual implicator

$$R(x,y) = \frac{R_C(x,y) + R_{T_{\mathbf{P}}}(x,y)}{2}.$$

By deresiduation, R generates the following quasi–copula

$$Q_R(x,y) = \min\left(\frac{x(\sqrt{1+8y}-1)}{2}, \frac{2xy}{1+\sqrt{x}}\right).$$

Notice that Q_R is a copula.

We expect that, sometimes, this kind of operation produces a useful tool to the construction of new copulas.

Acknowledgements. The third author was supported by the grant VEGA 1/3006/06 and MSM 619 8898 701. The fourth author gratefully acknowledges the support of the project "Metodi Stocastici in Finanza Matematica" of the Italian M.I.U.R..

References

1. C. Alsina, M.J. Frank, B. Schweizer, Associative functions. Triangular norms and copulas, World Scientific Publishing, Hackensack, NJ, 2006.
2. B. De Baets, R. Mesiar, Residual implicators of continuous t–norms, Proc. EUFIT 96, Aachen, 1996, pp. 37–41.
3. F. Durante, C. Sempi, Semicopulæ, Kybernetika 41 (2005) 315–328.
4. J. Fodor, T. Keresztfalvi, Nonstandard conjunctions and implications in fuzzy logic, Internat. J. Approx. Reason. 12 (1995) 69-84.
5. S. Gottwald, A Treatise on Many–Valued Logic, Studies in Logic and Computation, Research Studies Press, Baldock, 2001.
6. S. Gottwald, P. Hájek, Triangular norm-based mathematical fuzzy logics, in: E.P. Klement, R. Mesiar (eds.), Logical, algebraic, analytic, and probabilistic aspects of triangular norms, Elsevier, Amsterdam, 2005, pp. 275–299.
7. C. Genest, J.J. Quesada Molina, J.A. Rodríguez Lallena, C. Sempi, A characterization of quasi-copulas, J. Multivariate Anal. 69 (1999) 193–205.
8. P. Hájek, Metamathematics of fuzzy logic, Kluwer Academic Publishers, Dordrecht, 1998.
9. E.P. Klement, R. Mesiar, E. Pap, Triangular norms, Kluwer Academic Publishers, Dordrecht, 2000.
10. K.C. Maes, B. De Baets, A contour view on uninorm properties. Kybernetika 42 (2006), 303–318.
11. K.C. Maes, B. De Baets, On the structure of left-continuous t–norms that have a continuous contour line. Fuzzy Sets and Systems, in press.
12. M. Miyakoshi, M. Shimbo, Solutions of composite fuzzy relational equations with triangular norms, Fuzzy Sets and Systems 16 (1985) 53–63.
13. R.B. Nelsen, Copulas and quasi-copulas: an introduction to their properties and applications, in: E.P. Klement, R. Mesiar (eds.), Logical, algebraic, analytic, and probabilistic aspects of triangular norms, Elsevier, Amsterdam, 2005, pp. 391–413.
14. R.B. Nelsen, An introduction to copulas, Springer, New York, 2006.
15. F. Suárez García and P. Gil Álvarez, Two families of fuzzy integrals, *Fuzzy Sets and Systems* **18**, 67–81 (1986).

Vagueness as Semantic Indeterminacy: Three Bridges Between Probability and Fuzzy Logic

Christian G. Fermüller

Technische Universität Wien
A-1040 Vienna, Austria
chrisf@logic.at

It has been pointed out frequently and convincingly that in reasoning with vague, imprecise and uncertain information one should clearly distinguish between the two orthogonal dimensions of *probability* (i.e., degrees of uncertainty) and *fuzziness* (i.e., degrees of truth or membership), respectively. (See, e.g., the preface of [11]). The following simple example, that we will re-examine from different perspectives in our talk, should serve as a reminder.

Example 1. Suppose we gamble by throwing two (ordinary, fair) dices. In any round we can thus score any integer value ≥ 2 and ≤ 12. Clearly, the truth of an assertion like

(U) I will score more that 7 in the next round

is uncertain, but not vague. On the other hand, if, after having scored 8 in the last round, I assert the statement

(V) I have scored a rather high value in the last round

then this assertion is vague and might be considered a paradigmatic example of a proposition to which we may want to assign an intermediary degree of truth in the sense of formal fuzzy logic. Note that the truth value ($\in [0,1]$) thus associated with (V) does not refer to probability; at least not in the sense in which the truth of statement (U) refers to probability.

Without intending to blur the distinction between probability and fuzziness, we will explore a certain 'theory of vagueness' (in the sense of [12, 3]) that entails firm, non-trivial connections between probabilities and degrees of truth in its underlying semantic machinery. More specifically, we refer to the concept of a *precisification space*, consisting of (classical) interpretations, intended as admissible precisifications or sharpenings of vague notions and propositions. This concept is used in the so-called supervaluationists account of vagueness [7, 12, 14] in order to render the slogan that 'truth is supertruth' (i.e., truth in all admissible precisifications) amenable to a formal treatment. However, we depart from supervaluationism by endowing precisification spaces with probability measures that are intended to model the respective plausibility of choosing a particular precisification. Moreover, we extend the basic set-up of supervaluationism by considering not only propositions, but also entities like runs of a program (i.e., traces of automata or Turing machines) as well as levels of descriptions of elements in fuzzy sets (in the sense of [2]) as possible target objects of corresponding precisifications.

Building partly upon previous work [5, 6, 4], we will describe three quite different topics in fuzzy logic (taken here in Zadeh's wide sense) and show how one may profit from bringing to bear the above mentioned 'extended supervaluation' point of view in these different scenarios.

1 Playing Giles's game over precisification spaces

In a series of important papers [8–10] Robin Giles provided an alternative semantic foundation for Łukasiewicz logic **Ł** and reasoning with vague predicates in general. The main feature of Giles's approach consists in separating (1) the analysis of logical connectives and quantifiers from (2) the interpretation of (vague) atomic assertions. To this aim one models the stepwise reduction of logically complex assertions to their atomic components (1) by way of Lorenzen-style dialogue rules [13] that regulate idealized debates between a proponent and an opponent of an assertion. As for (2), Giles refers to bets on the results of 'dispersive elementary experiments' associated to the atomic statements that have been asserted by the dialogue participants.

As explained in [6] we modify Giles's original setting to evaluate atomic assertions with respect to precisification spaces, which are endowed with a probability measure μ on the set of precisification points W; μ is intended to model *degrees of plausibility* of the individual classical precisifications. This allows us to provide a dialogue game based semantics for Łukasiewicz logic enriched with an additional operator S that represents the modality '*It is supertrue that ...*' of supervaluation. The modal operator S behaves like $\Box$ in classical **S5**, and yet the resulting generalization S**Ł** of **Ł** is different from similar modal extensions of **Ł**(as e.g. described in chapter 8 of [11]). We argue that S**Ł** and its associated dialogue game serves as a formal model of the informal idea that logical evaluation of vague assertions consists in an *indeterministic evaluation* over corresponding admissible precisifications. It remains to be seen whether similar combinations of supervaluation and other t-norm based fuzzy logics can be achieved building on the dialogue games outlined in [1] and [5].

2 Computational complexity of vague descriptions

An approach that is closely related in spirit to the outlined generalization of supervaluation may help to adequately address what we have called 'the enigma of quantifying vague information' in [4]. The mentioned enigma consists in the fact that the intended *reduction* of descriptional complexity that often motivates the shift from an exact to a vague description of a given object or of a state of a system usually has to be payed for with a considerable *increase* in the computational complexity at the meta-level of formal representations of the descriptions in question. The fact that syntactic representations of fuzzy sets and propositions are more complex objects than crisp ones (within standard frameworks) is straightforwardly formalized using *Kolmogorov complexity*. E.g., referring to the example of throwing two dices, above, it should be clear that a formal description of the informational content of an assertion like (U) requires less resources in the sense of Kolmogorov complexity, in general, than a complete formal description of the contents of a vague statement like (V).

The question thus arises whether computational models of complexity, similar to the ones of Kolmogorov, Chaitin *et al.*, can be employed to formally represent the intended *decrease* of descriptional complexity in vague descriptions compared to crisp ones. We argue that probabilistic Turing machines (or similar devices) can benefitly be applied in this context. Note that the involved probability is *not* replacing the inherent vagueness, here, but is rather used to reflect the idea that vague descriptions may relate to precise descriptions in a way that is analogous to the relation between vague propositions and crisp propositions outlined in Section 1: it can be seen as indeterministically choosing precisifications while respecting different pre-assigned levels of plausibilities. Again, one may interpret the corresponding technical machinery as a concrete model of the overall idea that vagueness consists in semantic indeterminacy.

Formally, we will show how a properly generalized version of Kolmogorov complexity allows to quantify gradual changes in computational complexity when relaxing the level of precision in formal description of (binary represented) objects. The results reflect the intended reduction of complexity rather well.

3 Fuzzy elements and gradual sets

Recently Dubois and Prade [2] introduced the notion of a 'fuzzy element' within fuzzy set theory. This is intended to underline Zadeh's distinction between fuzziness and imprecision: "A fuzzy element is as precise as an element, just more gradual that the latter" (cited from [2]). As pointed out by Libor Behounek [unpublished manuscript] this concept is at stark variance with the standard approach to t-norm based fuzzy logic as developed by Hajek [11], Godo, Esteva, Montagna, and many others. Thus it is a major challenge to integrate these new concept into a theory of vagueness that remains close to the principles of ('Hajek style') deductive fuzzy logic.

We argue that, once more, the idea of interpreting vagueness as semantic indeterminacy, using probabilistically constrained evaluations in (classical) precisifications provides a key for relating the different approaches to fuzzy set theory that are relevant in this context. More exactly, we propose to identify the 'description level' λ, i.e., the argument of the function a_e representing a fuzzy element e, with the probability of choosing a particular precisification of the co-domain S of a_e. Interpreted in this way, a fuzzy element emerges as a collection of crisp elements endowed with a probability measure that models the respective plausibility λ of choosing the corresponding crisp element $a_e(\lambda)$ as de-fuzzificated counterpart of e.

References

1. A. Ciabattoni, C.G. Fermüller, G. Metcalfe: Uniform Rules and Dialogue Games for Fuzzy Logics. In: *LPAR 2004*, F. Baader, A. Voronkov (eds.), Springer LNAI 3452 (2005), 496-510.
2. D. Dubois, H. Prade: Fuzzy elements in a fuzzy set. In: *Proc. 10th Intern. Fuzzy Systems Assoc. (IFSA) Congress*, Beijing, 2005, Springer, to appear.
3. C.G. Fermüller: *Theories of Vagueness Versus Fuzzy Logic: Can Logicians Learn from Philosophers?* Neural Network World Journal, volume 13(5), 2003, 455-466.
4. C.G. Fermüller: The Enigma of Quantifying Vague Information. Abstract in: *The Logic of Soft Computing IV & Fourth Workshop of the ERCIM Working Group on Soft Computing*. Ed. V. Novak, M. Stepnicka. Research Report No. 77, University of Ostrava, Czech Republic, October 2005.
5. C.G. Fermüller: Revisiting Giles's Game – Reconciling Fuzzy Logic and Supervaluation. *Logic, Games and Philosophy: Foundational Perspectives*, Prague Colloquium October 2004, to appear.
6. C.G. Fermüller, R. Kosik: Combining supervaluation and degree based reasoning under vagueness. In: *LPAR 2006*, M. Hermann, A. Voronkov (eds.), Springer LNAI 4246 (2006), 212-226.
7. K. Fine: Vagueness, Truth and Logic, *Synthése* 30, 265-300, 1975.
8. R. Giles: A non-classical logic for physics. In: R. Wojcicki, G. Malinkowski (Eds.) *Selected Papers on Łukasiewicz Sentential Calculi*. Polish Academy of Sciences, 1977, 13-51.
9. R. Giles: A non-classical logic for physics. *Studia Logica* 33, vol. 4, (1974), 399-417.
10. R. Giles: Semantics for Fuzzy Reasoning. *Int. J. Man-Machine Studies* 17 (1982), 401-415.
11. P. Hájek: *Metamathematics of Fuzzy Logic*. Kluwer, 1998.
12. R. Keefe: *Theories of Vagueness*, Cambridge University Press, 2000.
13. P. Lorenzen: Logik und Agon. In: *Atti Congr. Internat. di Filosofia*, Vol. 4, Florence 1960, 187-194.
14. A. Varzi: Vagueness, Logic, and Ontology, *The Dialogue 1*, 135-154.

Level Dependent Capacities and the Generalized Choquet Integral

Silvio Giove[1], Salvatore Greco[2], and Benedetto Matarazzo[2]

[1] Department of Applied Mathematics, University Ca' Foscari of Venice,
Dorsoduro n. 3825/E - 30125 Venice, Italy
sgiove@unive.it

[2] Faculty of Economics, University of Catania Corso Italia 55, Catania 95129, Italy
{salgreco|matarazz}@unict.it

Abstract. We present a generalization of Choquet integral in which the capacity depends on the level of the aggregated variables (level dependent capacities). We show that as particular cases of our generalization of Choquet integral there are the Sugeno integral, the Sipos integral and the Cumulative Prospect Theory functional. We show also that many concepts such as Mobius transform, importance index, interaction index, k-order capacities and OWA operators, introduced in the research about Choquet integral can be generalized in the considered context.

In many decision problems a set of actions are evaluated with respect to a set of points of view called criteria. For example, in evaluating a car one can consider criteria such as maximum speed, price, acceleration, fuel consumption. In general, evaluations with respect to different criteria can be discordant. For example, very often when a car has a good maximum speed, it has also a high price and a high fuel consumption. Thus, in order to express a decision such as a choice among a set of car, it is necessary to aggregate the evaluations on considered criteria. This is the domain of multiple criteria decision analysis and in this context several methodologies have been proposed (for an extensive state-of-art surveys see [3]). One of the simplest aggregation procedures is the weighted sum of the evaluation of criteria. Some more complex aggregation procedures have proposed to take into account specific aspects in evaluating importance of criteria, such as interaction between criteria. The interaction of criteria has been considered through non-additive integrals such as Choquet integral [1] and Sugeno integral [2] (for a comprehensive survey see [3]). In this context, the importance of a set of criteria is not necessarily the sum of the importance of each criterion in the set. It can be greater or smaller, due to redundancy or synergy between criteria. An example of redundancy is the case of maximum speed and acceleration in evaluating cars: in fact a car that is speed very often has also a good acceleration and therefore, even if these two criteria can be very important for a person liking sport cars, the importance of them is smaller than the importance of the two criteria considered separately. An example of synergy is the case of maximum speed and price in evaluating cars: in fact a car that is speed, is often also a highly priced, and therefore a car with a high maximum speed and a not so high price is very well appreciated. So, the importance of these two criteria is greater than the importance of the two criteria considered separately. Introduction of a neutral point and consequent positive and negative sign of the evaluations with respect to considered criteria has been taken into account through some generalizations of the Choquet integral such as Sipos integral and Cumulative Prospect Theory functional. For example, in evaluating a car with respect to price, comfort and fuel consumption, for cars with low levels of price, comfort and fuel consumption, the price is more important than the comfort, while for cars having higher levels of price, comfort and fuel consumption, the comfort become more important than the fuel consumption. To model representation of the preferences in these situations we propose to assign importance to criteria depending on the level of evaluations and a consequent generalization of the Choquet integral [6]. This is a very flexible model to aggregate evaluations of set of criteria. Moreover this model presents as specific cases the Sugeno integral, the Sipos integral and the Cumulative Prospect Theory functional.

In this paper the generalization of the Choquet integral proposed by us takes into account the fact that the importance of criteria depends on the level of their evaluations. The starting point is the new concept of *Generalized Capacity* μ^G, that is a capacity defined over the criteria set, which depends on a parameter $t \in [0,1]$. If a continuity condition is satisfied, the *generalized Choquet* integral (GCI for brevity) with respect to the generalized capacity μ^G is given by:

$$\int_0^1 \mu^G(A(x,t),t)dt \tag{1}$$

where $A(x,t) = \{i \in N : x_i \geq t\}$.

A first result is a Theorem that states that if an aggregation function is monotonic, idempotent and tail independent, then it exists a Generalized Capacity so that the aggregation function can be expressed as the GCI with respect to this capacity. From this Theorem some other results can be obtained, the most important of them is the fact that the Sugeno integral too can be expressed as a GCI. The same can be obtained also for the Sipos integral and for the Cumulative Prospect Theory. Then, the main result of this contribution affirms that the Sugeno integral, the (classical) Choquet integral, the Sipos integral, and the Cumulative Prospect aggregating function can be expressed as particular case of the GCI. Moreover, also other aggregation functions, that cannot be expressed in none of the above operators (classical Choquet, Sipos and Sugeno integral, Cumulative Prospect Theory) can be expressed but a suitable GCI. Those are very surprising results, showing the very interesting generalization properties of the GCI.
As a consequence of the definition, we showed even that the usual items like the Mobius transform, the importance index, the interaction index, the k-order capacities and the OWA operators can be generalized in the considered context.

Acknowledgements

The research has been supported by the Italian Ministry of Education, University and Scientific Research (MIUR).

References

1. CHOQUET, G., Theory of Capacities, *Annales de l'Institut Fourier*, 5, 1954, 131-295.
2. SUGENO, M., *Theory of fuzzy integrals and its applications*, PhD thesis, Tokyo Institute of Technology, 1974.
3. GRABISCH, M., LABREUSCHE, C., *Fuzzy Measures and Integrals in MCDA*, in: J. Figueira, S. Greco, M. Ehrgott (eds.), Multiple Criteria Decision Analysis, Springer, New York, 2005, pp. 563-608.
4. FIGUEIRA, J., GRECO, S., EHRGOTT, M. (EDS.), *Multiple Criteria Decision Analsysis*, Springer, New York, 2005.
5. GREEN, J. R., JULLIEN, B., *Ordinal independence in nonlinear utility theory*,Journal of Risk and Uncertainty, 1 (1988) 355-387.
6. GRECO S., GIOVE S., MATARAZZO B., *Generalizing Choquet integral with level dependent capacities*, Proceedings of NOLTA2006, Bologna, (2006), CD-ROM.

Estimation of a Simple Linear Regression Model for Fuzzy Random Variables

Gil González-Rodríguez, Ángela Blanco, and M. Asunción Lubiano

Dpto. de Estadística e I.O. y D.M.
Universidad de Oviedo 33007 Oviedo, Spain
{gil|blancoangela.uo|lubiano}@uniovi.es

1 Introduction

The linear regression problem with fuzzy data has been previously treated in the literature from different points of view and by considering different kinds of input/output data (see, for instance, [6], [2] for recent surveys on the topic). In this work, a generalized simple linear regression statistical/probabilistic model in which both input and output data can be fuzzy subsets of $\mathbb{R}^p$ will be dealt with. Specifically, the least squares estimation problem will be addressed.

In Section 2 some notations and preliminary concepts will be presented. The considered Simple Linear Regression Model is stated in Section 3, and the least squares estimators are also obtained in it.

2 Preliminaries

Let $\mathcal{K}_c(\mathbb{R}^p)$ be the class of the nonempty compact convex subsets of $\mathbb{R}^p$ and let $\mathcal{F}_c(\mathbb{R}^p)$ be the *class of fuzzy sets* $U : \mathbb{R}^p \to [0,1]$ whose α-levels $U_\alpha \in \mathcal{K}_c(\mathbb{R}^p)$ for all $\alpha \in [0,1]$, where $U_0 = \mathrm{cl}(\{x \in \mathbb{R}^p \,|\, U(x) > 0\})$. Zadeh's extension principle [8] allows us to endow the space $\mathcal{F}_c(\mathbb{R}^p)$ with a sum and a product by a scalar extending the Minkowski sum and the product by a scalar on $\mathcal{K}_c(\mathbb{R}^p)$, since $(U+V)_\alpha = U_\alpha + V_\alpha$ and $(\lambda U)_\alpha = \lambda U_\alpha$ for all $U,V \in \mathcal{F}_c(\mathbb{R}^p)$, $\lambda \in \mathbb{R}$ and $\alpha \in [0,1]$ (with $A+B = \{a+b \,|\, a \in A,\, b \in B\}$ and $\lambda A = \{\lambda a \,|\, a \in A\}$ for $A,B \in \mathcal{K}_c(\mathbb{R}^p)$ and $\lambda \in \mathbb{R}$). Since $(\mathcal{F}_c(\mathbb{R}^p), +, \cdot)$ is not a linear space, it is often useful to consider the *Hukuhara difference* $U -_H V$, which is defined (if it exists) as the element $W \in \mathcal{F}_c(\mathbb{R}^p)$ so that $U = V + W$.

Let $\mathbb{S}^{p-1}$ be the unit sphere in $\mathbb{R}^p$ and $\langle \cdot, \cdot \rangle$ the inner product. The space $\mathcal{F}_c(\mathbb{R}^p)$ can be embedded onto the cone of the class of the Lebesgue integrable functions $\mathcal{L}(\mathbb{S}^{p-1} \times [0,1])$ by means of the mapping $s : \mathcal{F}_c(\mathbb{R}^p) \to \mathcal{L}(\mathbb{S}^{p-1} \times [0,1])$ which associates each $U \in \mathcal{F}_c(\mathbb{R}^p)$ into its *support function* $s(U) = s_U : \mathbb{S}^{p-1} \times [0,1] \to \mathbb{R}$, with $s_U(u,\alpha) = \sup_{w \in U_\alpha} \langle u, w \rangle$ for any $u \in \mathbb{S}^{p-1}$ and $\alpha \in [0,1]$. The support function is semilinear, in the sense that, $s_{U+V} = s_U + s_V$ and $s_{\lambda U} = \lambda s_U$ if $\lambda \geq 0$. Furthermore, if $U -_H V$ exists, then $s_{U -_H V} = s_U - s_V$ for all $U,V \in \mathcal{K}_c(\mathbb{R}^p)$ (see, for instance, [3]).

The least squares method to be considered will be based on the *generalized metric* D_K by Körner and Näther [4] on $\mathcal{F}_c(\mathbb{R}^p)$, which $U,V \in \mathcal{F}_c(\mathbb{R}^p)$ is defined as follows:

$$[D_K(U,V)]^2 = \int_{(\mathbb{S}^{p-1})^2 \times [0,1]^2} \big(s_U(u,\alpha) - s_V(u,\alpha)\big)\big(s_U(v,\beta) - s_V(v,\beta)\big)\, dK(u,\alpha,v,\beta),$$

where K is a positive definite and symmetric kernel; D_K coincides with a generic L_2 distance w.r.t. K on the Banach space $\mathcal{L}(\mathbb{S}^{p-1} \times [0,1])$. Thus, if we denote by $< \cdot, \cdot >_K$ the corresponding inner product, we have that $D_K(U,V) = < s_U - s_V, s_U - s_V >_K$.

A *Fuzzy Random Variable* (FRV) associated with a probability space $(\Omega, \mathcal{A}, P)$ in Puri and Ralescu's sense (see [7]) is a mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^p)$ such that the α-level mappings $\mathcal{X}_\alpha : \Omega \to \mathcal{K}_c(\mathbb{R}^p)$ are random sets for all $\alpha \in [0,1]$. Alternatively, an FRV is an $\mathcal{F}_c(\mathbb{R}^p)$-valued random element (i.e. a Borel-measurable mapping) when the D_K-metric is considered on $\mathcal{F}_c(\mathbb{R}^p)$ (see [1] and [4]), whence the probability distribution associated with the FRV and the independence can be formalized as usual.

Let $\mathcal{X}$ be a FRV such that $|\mathcal{X}_0| \in L^1(\Omega,\mathcal{A},P)$ (with $|X|(\omega) = \sup\{|x| \,|\, x \in X(\omega)\}$ for all $\omega \in \Omega$). Then, the *expected value (or mean)* of $\mathcal{X}$ is the unique $E(\mathcal{X}) \in \mathcal{F}_c(\mathbb{R}^p)$ such that $\big(E(\mathcal{X})\big)_\alpha =$ Aumman's integral of the random set $\mathcal{X}_\alpha$ for all $\alpha \in [0,1]$, that is,

$$\big(E(\mathcal{X})\big)_\alpha = \big\{E(f|P) \big| f : \Omega \to \mathbb{R}^p, f \in L^1, f \in \mathcal{X}_\alpha \, a.s.[P]\big\}$$

(see [7]). If, moreover, $E(|\mathcal{X}|^2) < \infty$, then, the *variance* of $\mathcal{X}$ is defined as $\mathrm{Var}(\mathcal{X}) = E\left([D_K(\mathcal{X},E[\mathcal{X}])]^2\right)$ (see, for instance, [5] and [4]).

Let $(\mathcal{X},\mathcal{Y}) : \Omega \to \mathcal{F}_c(\mathbb{R}^p) \times \mathcal{F}_c(\mathbb{R}^p)$ be a two-dimensional fuzzy random variable, and consider a random sample $\{\mathcal{X}_i,\mathcal{Y}_i\}_{i=1}^n$ obtained from $(\mathcal{X},\mathcal{Y})$. $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ will denote the sample means (that is, $\overline{\mathcal{X}} = (\mathcal{X}_1 + \ldots + \mathcal{X}_n)/n$ and $\overline{\mathcal{Y}} = (\mathcal{Y}_1 + \ldots + \mathcal{Y}_n)/n$), $\hat{\sigma}^2_{\mathcal{X}}$ will denote the sample variance (that is, $\hat{\sigma}^2_{\mathcal{X}} = \overline{D_K(\mathcal{X},\overline{\mathcal{X}})^2}$), and $\hat{\sigma}^2_{\mathcal{X},\mathcal{Y}}$ will denote the sample covariance (that is, $\hat{\sigma}^2_{\mathcal{X},\mathcal{Y}} = \overline{< s_{\mathcal{Y}} - s_{\overline{\mathcal{Y}}}, s_{\mathcal{X}} - s_{\overline{\mathcal{X}}} >_K}$).

3 Least squares estimators of a simple linear regression model between FRVs

The Simple Linear Regression Model is to be considered is $\mathcal{Y} = a\mathcal{X} + \varepsilon_{\mathcal{X}}$, where $a \in \mathbb{R}$ and $\varepsilon_{\mathcal{X}}$ is a fuzzy random variable with expected value $E[\varepsilon_{\mathcal{X}}] = B \in \mathcal{F}_c(\mathbb{R}^p)$, which implies that $E[\mathcal{Y}|\tilde{x}] = a\tilde{x} + B$ for any $\tilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$.

An alternative model would be given by $\mathcal{Y} = a\mathcal{X} + B + \varepsilon_{\mathcal{X}}$, with $E[\varepsilon_{\mathcal{X}}] = \chi_{\{0\}} \in \mathcal{F}_c(\mathbb{R}^p)$ (where $\chi_{\cdot}$ stands for the characteristic function of a classical set). However, in this case the lack of linearity of $\mathcal{F}_c(\mathbb{R}^p)$ would imply the errors to be degenerated into random variables, although the regression function is also $E[\mathcal{Y}|\tilde{x}] = a\tilde{x} + B$. Thus, in order to consider fuzzy-valued errors, the independent term would be included in the formalization of the possible errors.

On the other hand, it should be noted that the model $\mathcal{Y} = a\mathcal{X} + \varepsilon_{\mathcal{X}}$ forces the existence of the Hukuhara difference $\mathcal{Y} -_H a\mathcal{X}$. We propose to take this fact into account in order to estimate the model and to look for a kind of *restricted Least Squares estimators*. That is, we constrain the estimator of a to the set $A = \{a^* \in \mathbb{R} \,|\, \mathcal{Y}_i -_H a^*\mathcal{X}_i$ exists for all $i = 1,\ldots,n\}$. Thus, the least squares problem consist of looking for $\widehat{a} \in \mathbb{R}$ and $\widehat{B} \in \mathcal{F}_c(\mathbb{R}^p)$ in order to

$$\begin{array}{l} \text{Minimize } \dfrac{1}{n}\displaystyle\sum_{i=1}^n D_K^2(\mathcal{Y}_i, a\mathcal{X}_i + B) \\ \text{subject to} \\ a \in A \end{array}$$

Since the variance w.r.t. the D_K metric satisfies the Fréchet approach (see [4]), if $a \in A$, then the minimum of $\frac{1}{n}\sum_{i=1}^n D_K^2(\mathcal{Y}_i, a\mathcal{X}_i + B)$ over $B \in \mathcal{F}_c(\mathbb{R}^p)$ is attained at $B(a) = \overline{\mathcal{Y}} -_H a\overline{\mathcal{X}}$. Consequently, the minimization problem reduces to

$$\begin{array}{l} \text{Find } \widehat{a} \in \mathbb{R} \text{ minimizing } \phi(a) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n D_K^2(\mathcal{Y}_i -_H a\mathcal{X}_i, \overline{\mathcal{Y}} -_H a\overline{\mathcal{X}}) \\ \text{subject to } a \in A. \end{array}$$

By applying properties of the support function and the inner product in $\mathcal{L}(\mathbb{S}^{p-1} \times [0,1])$, we have that the objective function $\phi(a)$ can be decomposed as follows:

$$\phi(a) = \frac{1}{n}\sum_{i=1}^n < s_{\mathcal{Y}_i} - s_{\overline{\mathcal{Y}}}, s_{\mathcal{Y}_i} - s_{\overline{\mathcal{Y}}} >_K + \frac{1}{n}\sum_{i=1}^n < s_{a\mathcal{X}_i} - s_{a\overline{\mathcal{X}}}, s_{a\mathcal{X}_i} - s_{a\overline{\mathcal{X}}} >_K$$

$$-2\frac{1}{n}\sum_{i=1}^{n} < s_{\mathcal{Y}_i} - s_{\overline{\mathcal{Y}}}, s_{aX_i} - s_{a\overline{X}} >_K .$$

On the other hand, it is possible to check that either $A = \mathbb{R}$, or there exists $a^0, b^0 \in [0, +\infty)$, so that $A = [-a^0, b^0]$. The homogeneity property of the support function w.r.t. the product by a scalar is only satisfied for positive constant, whence positive and negative parts of A will be treated separately. For the sake of simplicity, the reasonings will be developed in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ by denoting $a^0 = b^0 = \infty$ in case $A = \mathbb{R}$.

Concerning the positive part of A, if $0 \le a \le b^0$, we have that

$$\phi(a) = \hat{\sigma}^2_{\mathcal{Y}} + a^2 \hat{\sigma}^2_{X} - 2a\hat{\sigma}_{X,\mathcal{Y}}.$$

This function is continuous, differentiable and convex, and then it is easy to find that the minimum is attained at $a_1 = \beta\frac{\hat{\sigma}_{X,\mathcal{Y}}}{\hat{\sigma}^2_X}$, whenever $\hat{\sigma}^2_X > 0$, where

$$\beta = \begin{cases} 0 & \text{if } \hat{\sigma}_{X,\mathcal{Y}} \le 0 \\ \min\left\{1, \frac{b^0}{\hat{\sigma}_{X,\mathcal{Y}}/\hat{\sigma}^2_X}\right\} & \text{if } \hat{\sigma}_{X,\mathcal{Y}} > 0 \end{cases}$$

Analogously, if $-a^0 \le a \le 0$, we have that

$$\phi(a) = \hat{\sigma}^2_{\mathcal{Y}} + a^2 \hat{\sigma}^2_{-X} + 2a\hat{\sigma}_{-X,\mathcal{Y}}.$$

It is easy to check that $\hat{\sigma}^2_{-X} = \hat{\sigma}^2_X$, due to the symmetry of the kernel K in the distance D_K. However, there is not a general relationship between $\hat{\sigma}_{-X,\mathcal{Y}}$ and $\hat{\sigma}_{X,\mathcal{Y}}$. Thus, the minimum of $\phi(a)$ in this case is attained at $a_2 = -\alpha\frac{\hat{\sigma}_{-X,\mathcal{Y}}}{\hat{\sigma}^2_X}$ whenever $\hat{\sigma}^2_X > 0$, where

$$\alpha = \begin{cases} 0 & \text{if } \hat{\sigma}_{-X,\mathcal{Y}} \le 0 \\ \min\left\{1, \frac{a^0}{\hat{\sigma}_{-X,\mathcal{Y}}/\hat{\sigma}^2_X}\right\} & \text{if } \hat{\sigma}_{-X,\mathcal{Y}} > 0 \end{cases}$$

The global minimum should be computed by taking into account the positive and the negative part. To simplify the computations in practice, it is possible to express the different cases that arise in terms of the sample covariances. Specifically, we can check that the minimum of $\phi(a)$ over $[-a^0, b^0] \subset \overline{\mathbb{R}}$ is attained at

$$a^* = \begin{cases} \beta\frac{\hat{\sigma}_{X,\mathcal{Y}}}{\hat{\sigma}^2_X} - \alpha\frac{\hat{\sigma}_{X,\mathcal{Y}}}{\hat{\sigma}^2_X} & \text{if } \alpha = 0 \text{ or } \beta = 0 \\ -\alpha\frac{\hat{\sigma}_{-X,\mathcal{Y}}}{\hat{\sigma}^2_X} & \text{if } \frac{\hat{\sigma}^2_{-X,\mathcal{Y}}}{\hat{\sigma}^2_{X,\mathcal{Y}}} \ge \frac{2\beta-\beta^2}{2\alpha-\alpha^2} \text{ and } \alpha\cdot\beta \ne 0 \\ \beta\frac{\hat{\sigma}_{X,\mathcal{Y}}}{\hat{\sigma}^2_X} & \text{if } \frac{\hat{\sigma}^2_{-X,\mathcal{Y}}}{\hat{\sigma}^2_{X,\mathcal{Y}}} \le \frac{2\beta-\beta^2}{2\alpha-\alpha^2} \text{ and } \alpha\cdot\beta \ne 0 \end{cases}$$

References

1. A. Colubi, J. S. Domínguez-Menchero, M. López-Díaz, and D. A. Ralescu. A $D_E[0,1]$-representation of random upper semicontinuous functions. *Proc. Amer. Math. Soc.*, 130:3237–3242, 2002.
2. R. Coppi, P. DUrso, P. Giordani, and A. Santoro Least squares estimation of a linear regression model with LR fuzzy response *Comp. Stat. Data Anal.*, 51:267–286, 2006.
3. P. Diamond and P. Kloeden. *Metric Spaces of Fuzzy Sets: Theory and Applications.* World Scientific, Singapore, 1994.
4. R. Körner and W. Näther. On the variance of random fuzzy variables. In C. Bertoluzza, M.A. Gil, and D.A. Ralescu, editors, *Statistical Modeling, Analysis and Management of Fuzzy Data*, pages 22–39, 2002.
5. M.A. Lubiano,M.A. Gil, M. López-Díaz, and M.T. López-García The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable *Fuzzy Sets and Systems*, 111:307–317, 2000.

6. W. Näther. Regression with fuzzy random data *Comp. Stat. Data Anal.*, 51:235–252, 2006.
7. M. L. Puri and D. A. Ralescu. Fuzzy random variables. *J. Math. Anal. Appl.*, 114:409–422, 1986.
8. L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, I. *Inform. Sci.*, 8:199–249, 1975.

On the Vertices of the K-Additive Core of Capacities

Michel Grabisch[1] and Pedro Miranda[2]

[1] Université Paris I - Panthéon-Sorbonne
106-112, Bd. de l'Hôpital, 75013 Paris, France
Michel.Grabisch@lip6.fr
[2] Universidad Complutense de Madrid
Plaza de Ciencias, 3, 28040 Madrid, Spain
pmiranda@mat.ucm.es

1 Introduction

The core of a capacity or a game is a fundamental concept, both in decision making theory and in cooperative game theory. In decision making, it is the set of probability measures which are coherent with the information given by the capacity in the representation of uncertainty [8]. In game theory, it is the set of imputations (additive games) that can be given to players so that no subcoalition of the grand coalition has interest to form.

The properties of the core are well known, most of them have been shown by Shapley [7]. In many cases, it happens that the core is empty. A sufficient and necessary condition for nonemptiness is known for capacities and games, which is called balancedness. In particular, convex capacities have a nonempty core.

Since having an empty core is not a favorable situation, either in decision making or in game theory, it may be an alternative solution to look for more general concepts. For example, since the core contains additive capacities or games, we may relax additivity to a weaker notion: k-additivity, proposed by Grabisch [2]. We may call this new notion the *k-additive core*.

Some studies on the k-additive core have already been done by the authors, see, e.g., [3, 6]. It happens that the structure of the k-additive core is much more complex that the one of the classical core. In particular, the set of its vertices is not known. The aim of this paper is to provide new insights in this direction. More details can be found in [4].

2 Background

Throughout the paper, we consider a finite universal set X, with $|X| = n$. We use indifferently 2^X or $\mathcal{P}(X)$ to denote the set of subsets of X, and the set of subsets of X containing at most k elements is denoted by $\mathcal{P}^k(X)$, while $\mathcal{P}^k_*(X) := \mathcal{P}^k(X) \setminus \{\emptyset\}$.

A *capacity* $\mu : 2^X \to \mathbb{R}_+$ is a function such that $\mu(\emptyset) = 0$, and $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$ (*monotonicity*). A capacity is *normalized* if $\mu(X) = 1$. We assume in this paper that capacities are normalized. The set of capacities on X is denoted by $\mathcal{FM}(X)$. A capacity μ on X is said to be *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$, *convex* if $\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B)$, for all $A, B \subseteq X$, and *k-monotone* for $k \geq 2$ if for any family of k subsets $A_1, \ldots A_k$, it holds

$$\mu(\bigcup_{i=1}^{k} A_i) \geq \sum_{\substack{K \subseteq \{1,\ldots,k\} \\ K \neq \emptyset}} (-1)^{|K|+1} \mu(\bigcap_{j \in K} A_j).$$

A capacity is *totally monotone* (belief function) if it is k-monotone for all $k \geq 2$. Convexity is equivalent to 2-monotonicity.

Let μ be a capacity on X. The *Möbius transform* of μ is a function $m : 2^X \to \mathbb{R}$ defined by:

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B).$$

A capacity μ is said to be *k-additive* for some integer $k \in \{1, \dots, n\}$ if $m(A) = 0$ whenever $|A| > k$, and there exists some A such that $|A| = k$, and $m(A) \neq 0$. The set of at most k-additive capacities on X is denoted by $\mathcal{FM}^k(X)$.

It is known from [1] that k-monotonicity is equivalent to

$$\sum_{A \subseteq L \subseteq B} m(L) \geq 0, \quad \forall A \subseteq B \subseteq X, \quad |A| \leq k.$$

The *core* of a capacity μ is defined by:

$$\mathcal{C}(\mu) := \{\nu \in \mathcal{FM}^1(X) \mid \nu(S) \geq \mu(S), \forall S \subseteq X\}.$$

A *maximal chain* in 2^X is a sequence of subsets $A_0 := \emptyset, A_1, \dots, A_{n-1}, A_n := X$ such that $A_i \subset A_{i+1}$, $i = 0, \dots, n-1$. The set of maximal chains of 2^X is denoted by $\mathcal{M}(2^X)$.

To each maximal chain $C := \{\emptyset, A_1, \dots, A_n = X\}$ in $\mathcal{M}(2^X)$ corresponds a unique permutation σ on X such that $A_1 = \sigma(1)$, $A_2 \setminus A_1 = \sigma(2)$, ..., $A_n \setminus A_{n-1} = \sigma(n)$. The set of all permutations over X is denoted by $\mathfrak{S}(X)$. Let μ be a capacity. To each permutation σ (or maximal chain C) we assign a *marginal worth vector* p^σ (or p^C) in $\mathbb{R}^n$ defined by:

$$p^C_{\sigma(i)} := \mu(A_i) - \mu(A_{i-1}).$$

Any marginal worth vector forms a probability distribution over X, and hence defines an additive capacity. The following is immediate.

Proposition 1. *Let μ be a capacity on X, and C a maximal chain of 2^X. Then $p^C(A) = \mu(A), \quad \forall A \in C$.*

Theorem 1. *The following are equivalent.*

(i) *μ is a convex capacity*
(ii) *all marginal worth vectors p^σ, $\sigma \in \mathfrak{S}(X)$ belong to the core of μ*
(iii) $\mathcal{C}(\mu) = \mathrm{co}(\{p^\sigma\}_{\sigma \in \mathfrak{S}(X)})$
(iv) $\mathrm{ext}(\mathcal{C}(\mu)) = \{p^\sigma\}_{\sigma \in \mathfrak{S}(X)}$,

where $\mathrm{co}(K)$ *and* $\mathrm{ext}(K)$ *denote respectively the convex hull and the extreme points of some convex set K.*

(i) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iv) are due to Shapley [7], while (ii) $\Rightarrow$ (i) was proved by Ichiishi [5].

3 Vertices of the k-additive core

Let μ be a capacity on X, and $1 \leq k \leq n-1$. The *k-additive core* of μ is defined by:

$$\mathcal{C}^k(\mu) := \{\nu \in \mathcal{FM}^k(X) \mid \nu(S) \geq \mu(S), \forall S \subseteq X\}.$$

Similarly, we introduce $\mathcal{BC}^k(\mu)$ the set of k-additive belief functions dominating μ. We put $N(k) := \binom{n}{1} + \cdots + \binom{n}{k}$. A first fact is the following.

Proposition 2. *For any capacity μ, $\mathcal{C}^k(\mu)$ and $\mathcal{BC}^k(\mu)$ are closed convex $(N(k)-1)$-dimensional polytopes.*

We denote by $\prec$ a total (strict) order on $\mathcal{P}^k_*(X)$, $\preceq$ denoting the corresponding large order. For any $B \in \mathcal{P}^k_*(X)$, we define

$$\mathcal{A}(B) := \{A \subseteq X \mid A \supseteq B, \forall K \subseteq A, K \in \mathcal{P}^k_*(X), K \preceq B\}$$

the *achievable family of B*. It is easy to see that $\{\mathcal{A}(B)\}_{B\in\mathcal{P}^k_*(X)}$ is a partition of $\mathcal{P}(X) \setminus \{\emptyset\}$.

A total order $\prec$ on $\mathcal{P}^k_*(X)$ is said to be *compatible* if for all $i, j \in X$, $i \prec j$ implies $S \cup i \prec S \cup j$, for any $S \in \mathcal{P}^{k-1}(X)$, $i, j \notin S$. It is said to be $\subseteq$*-compatible* if $A \subseteq B$ implies $A \prec B$. Lastly, $\prec$ is said to be *strongly compatible* if it is compatible and $\subseteq$-compatible, and *weakly compatible* if only compatible.

Proposition 3. *Assume $\prec$ is compatible. For any $B \in \mathcal{P}^k_*(X)$ such that $\mathcal{A}(B) \neq \emptyset$, $\mathcal{A}(B)$ is a Boolean lattice with bottom element B. The top element is denoted by $\check{B}$.*

$\subseteq$-compatibility is a sufficient and necessary condition for the nonemptiness of all achievable families.

Let μ be a capacity on X, m its Möbius transform, and $\prec$ some total order on $\mathcal{P}^k_*(X)$. We define $\mu_\prec$ by its Möbius transforms as follows:

$$m_\prec(B) := \begin{cases} \sum_{A\in\mathcal{A}(B)} m(A), & \text{if } \mathcal{A}(B) \neq \emptyset \\ 0, & \text{else} \end{cases} \qquad (1)$$

for all $B \in \mathcal{P}^k_*(X)$, $m_\prec(\emptyset) := 0$. Since achievable families form a partition of 2^X, $m_\prec$ satisfies $\sum_{B\subseteq X} m_\prec(X) = 1$, hence $\mu_\prec(X) = 1$. The following can be shown.

Proposition 4. *If $\prec$ is compatible, then for any nonempty achievable family $\mathcal{A}(B)$, $\mu_\prec(\check{B}) = \mu(\check{B})$.*

(see the analogy with Prop. 1).

Proposition 5. *Let μ be a capacity on X. Then $\mu_\prec$ is a belief function for any compatible order $\prec$ if and only if μ is k-monotone.*

The following propositions are analogous to the Shapley-Ichiishi theorem above.

Proposition 6. *Let μ be a capacity on X. Then $\mu_\prec \in \mathcal{C}^k(\mu)$ for all compatible orders $\prec$ if and only if μ is $(k+1)$-monotone.*

Proposition 7. *Let μ be a $(k+1)$-monotone capacity. Then*

(i) If $\prec$ is strongly compatible, then $\mu_\prec$ is a vertex of $\mathcal{C}^k(\mu)$.
(ii) If $\prec$ is compatible, then $\mu_\prec$ is a vertex of $\mathcal{BC}^k(\mu)$.

However, there are many vertices that are not belief functions. Experiments conducted with the PORTA software finding vertices and facets of polyhedra show that, for example, the set of vertices of $\mathcal{C}^2(\mu)$ of the following 3-monotone capacity μ with $n = 3$

A	1	2	3	12	13	23	123
$m(A)$	0	0.1	0.2	0.1	0	0.2	0.4
$\mu(A)$	0	0.1	0.2	0.2	0.2	0.5	1

has 48 elements, whose only 3 are belief functions.

Let us examine more precisely the vertices induced by strongly compatible orders. In fact, there are much fewer than expected, since many strongly compatible orders lead to the same $\mu_\prec$ (hence the experimental result above). We can show the following.

Proposition 8. *The number of vertices of $\mathcal{C}^k(\mu)$ given by strongly compatible orders is at most $\frac{n!}{k!}$.*

We examine the case of compatible orders being not strongly compatible (let us call them *weakly compatible*).

Proposition 9. *Suppose μ is a $(k+1)$-monotone capacity which satisfies $\mu(\{i\}) > 0$ for all $i \in X$. Then no weakly compatible order can produce a vertex of $\mathcal{C}^k(\mu)$.*

The following result seems to be true.

Conjecture 1. Assume $k < n$ and μ is a $(k+1)$-monotone capacity such that $\mu(\{i\}) > 0$ for all $i \in N$. Then μ^* is a vertex of $\mathcal{C}^k(\mu)$ such that $m^* \geq 0$ if and only if there exists a strongly compatible order $\prec$ such that $m^* = m_{\prec}$.

References

1. A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.
2. M. Grabisch. *k*-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.
3. M. Grabisch. On lower and upper approximation of fuzzy measures by *k*-order additive measures. In *7th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98)*, pages 1577–1584, Paris, France, July 1998.
4. M. Grabisch and P. Miranda. A study of the *k*-additive core of capacities through achievable families. In *SCIS-ISIS 2006, 3nd Int. Conf. on Soft Computing and Intelligent Systems and 7th Int. Symp. on Advanced Intelligent Systems*, Yokohama, Japan, September 2006.
5. T. Ichiishi. Super-modularity: applications to convex games and to the greedy algorithm for LP. *J. Econom. Theory*, 25:283–286, 1981.
6. P. Miranda, M. Grabisch, and P. Gil. Dominance of capacities by *k*-additive belief functions. *Eur. J. of Operational Research*, 175:912–930, 2006.
7. L. S. Shapley. Core of convex games. *Int. J. Game Theory*, 1:11–26, 1971.
8. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

On Fuzzy Logics of "Probably" — a Survey

Petr Hájek

Institute of Computer Science
Academy of Sciences of the Czech Republic
182 07 Prague, Czech Republic
hajek@cs.cas.cz

The distinction between fuzzy logic in broad and narrow sense as well between fuzzy logic and probability theory is stressed; building bridges is wanted. We survey one of possibilities - the logic of the fuzzy notion on being probable. Another possibility, the theory of states on MV-algebras is not surveyed, but the notion of a state is assumed to be known, see e.g. [16, 12]. Also the basic fuzzy logic BL and related logics (Łukasiewicz, Gödel, product logic and their combinations) are assumed known. We survey the existing literature and add some few new results.

Propositional logic. Let L_1 be the Boolean logic or a fuzzy logic, L_2 a fuzzy logic. The fuzzy probability logic $FP(L_1, L_2)$ is a fuzzy modal logic whose non-modal formulas are those of L_1, atomic modal formulas have the form $\mathbf{P}\varphi$ where $\mathbf{P}$ is a modality read "probably" and other modal formulas are formed from atomic ones by means of the logic L_2. Probabilistic Kripke models have the form (W, e, σ) where W is a non-empty (at most countable) set of possible worlds, e evaluates in each possible world all atomic formulas by truth values from the standard set of truth values of L_1 and $\sigma : W \to [0,1]$ with $\sum \sigma(w) = 1$. $\|\mathbf{P}\varphi\| = \sum e_{L_1}(\varphi, w)\sigma(w)$ is the probability of φ. The problems concerning this logic include axiomatization, completeness, computational complexity. Several particular cases were studied.

(1) $FP(Bool,Ł)$, $FP(Bool,Ł\Pi)$, $FP(Bool,RPL)$ ([10, 9]). The axioms are Boolean axioms for non-modal formulas, axioms of the fuzzy logic in question for modal formulas and three axioms $(FP1)-(FP3)$ for the modality. (Deduction rules modus ponens and generalization for $\mathbf{P}$.) The first two logics have finite strong (standard) completeness, the last one has Pavelka completeness (degree of provability equals degree of truth). The set $Sat(FP(Bool,Ł))$ of satisfiable formulas is NP-complete, $Sat(FP(Bool,Ł\Pi))$ is in $PSPACE$ [11]. Also the set $Taut(FP(Bool,Ł))$ is co-NP complete, $Taut(FP(Bool,Ł\Pi))$ is $PSPACE$.

(2) $FP(Ł_n,Ł)$, $FP(Ł_n,Ł\Pi)$, $FP(Ł_n,RPL)$ where $Ł_n$ is the logic of the standard Łukasiewicz (MV) algebra with n elements. These logics are studied in [3]; the authors investigate double semantics, that with probabilistic Kripke models and a more general semantics interpreting the modality $\mathbf{P}$ using the notion of a state; the axiomatic systems (analogous to those of logics in (1)) are complete similarly to the situation in (1). Analyzing [11] I can show the following uniform complexity result: the set $\{(\Phi, n) \mid \Phi \text{ is } FP(Ł_n,Ł)\text{-satisfiable}\}$ is NP-complete. [1] Similarly for L_2 being $Ł\Pi$ and the complexity being $PSPACE$.

(3) $FP(G,Ł)$ where G stands for Gödel logic. The set of satisfiable formulas is NPcomplete - proof by reducing to the uniform complexity result for $FP(Ł_n,Ł)$ above. Problem: find an axiomatization; axioms for the modality above do not work. (But e.g. $P\neg\varphi \vee P\neg\neg\varphi$ is a tautology.)

(4) $FP(Ł,RPL)$. This logic has Pavelka completeness (mentioned in [3]). No results on computational complexity yet.

[1] Caution: the details of the proof have still to be elaborated.

(5) Let us mention three logics dealing with conditional provability in our fuzzy style: $FCP(Bool, \text{Ł}\Pi)$ [3, 8], $FP(SL\Pi)$ [4] and $FP_k(RPL_\Delta)$ [2].

Predicate calculus. The approach from [9] will be critically surveyed and slightly simplified. The logics can be called $FP(\text{Ł}\forall$ and $FP(RPL\forall)$; the probabilistic semantics works with models $(M, (r_Q)_Q, \sigma)$ where $(M, (r_Q)_Q)$ is an usual *standard* countable interpretation of a given predicate language and σ is as above (with the domain M instead of a W) and the language is expanded by a new *quantifier* **P**: if φ is a formula then so is $(\mathbf{P}x)\varphi$ and $\|(\mathbf{P}x)\varphi(x,\dots)\|_{\mathbf{M},v} = \sum_{a\in M}\|\varphi(a,\dots)\|_{\mathbf{M},v}\cdot\sigma(a)$. An alternative semantics replaces the σ by a p-state (state satisfying some additional conditions). Each model in the sense of the probabilistic semantics defines a p-state whose domain is the set of all parametrically definable fuzzy subsets of M. The axioms for the new quantifier are

(μ1) $(Px)\nu \equiv \nu$ x not free in ν
(μ2) $(Px)\neg\varphi \equiv \neg(Px)\varphi$
(μ3) $(Px)(\varphi \to \psi) \to ((Px)\varphi \to (Px)\psi)$
(μ4) $(Px)(\varphi \oplus \psi) \equiv ([(Px)\varphi \to (Px)(\varphi \& \psi)] \to (Px)\psi)$
(μ5) $(\forall x)\varphi \to (Px)\varphi$
(μ6) $(Px)(Py))\varphi \equiv (Py)(Px)\varphi$

The logic $FP(RPL\forall)$ has strong standard Pavelka completeness with respect to the alternative semantics; it follows that the set of (alternative) tautologies of this logic is a Π_2-complete set in the sense of the arithmetical hierarchy.

There are several open problems, among them: Axiomatizability and decidability of $FP(\text{Ł},\text{Ł})$? (See [3] page 19.) Further, completeness (Pavelka) of $FP(RPL\forall)$ w.r.t. probabilistic models.

References

1. Fagin R., Halpern J.Y., Megido N.: A logic for reasoning about probabilities. Information and Computation 87 (1990) 78-128
2. Flaminio T.: A zero-layer based fuzzy probabilistic logic for cnditional probability. Preprint.
3. Flaminio T., Godo L: A logic for reasoning about probability of fuzzy events. To appear in Fuzzy Sets and Systems.
4. Flaminio F., Montagna F.: A logical and probabilistic treatment on conditinal probability. Preprint.
5. Godo L., Esteva F., Hájek Petr: Reasoning about Probability usIing Fuzzy Logic. godoeh* - Neural Network World, Vol. 10, 2000, No. 5, pp. 811-824
6. Godo L., Esteva F., Hájek P. A Fuzzy Modal Logic for Belief Functions. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence. Vol. : 1. - Seattle, Morgan Kaufmann 2001, pp. 723-729
7. Godo L., Hájek P., Esteva F.: A Fuzzy Modal Logic for Belief Functions. - Fundamenta Informaticae, Vol. 57, 2003, pp. 127-146
8. Godo L., Marchioni E.: Coherent conditional probability in a fuzzy logic setting. To appear in Logic J. of the IGPL.
9. Hájek P.: Metamathematics of fuzzy logic. Kluwer 1998
10. Hájek P., Godo L., Esteva F.: Fuzzy Logic and Probability. In: Uncertainty in Artificial Intelligence. Proceedings of the 11. Conference. (Ed.: Besnard P., Hanks S.) - San Francisco, Morgan Kaufmann 1995, pp. 237-244 Held: UAI95, Montreal, CA, 95.08.18-95.08.20
11. Hájek P., Tulipani S.: Complexity of Fuzzy Probability Logic. - Fundamenta Informaticae, Vol. 45, 2001, pp. 207-213
12. Kroupa T.: States and conditinal probability on MV-algebras (dissertation)
13. Marchioni E.: Possibilistic conditioning framed in fuzzy logics. Int.J. Aprox. reasoning (in press)
14. Marchioni E., Godo L.: A logic for reasoning about coherent conditional probabilities: a fuzzy modal logic approach. JELIA 04 Lisbon, Lect. Notes in AI 3229, 213-225
15. Mertanen J., Turunen E.: Problematic extensions of states o BL-algebras, preprint
16. Mundici D.: Bookmaking over infinite-valued events. Int. J. of Approximate Reasoning (in press)

Limit Points of Sequences of Fuzzy Real Numbers

Jaroslav Hančl[1], Ladislav Mišík[1] and János T. Tóth[1]

Institute for Research and Applications of Fuzzy Modeling,
University of Ostrava,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
{hancl|ladislav.misik|janos.toth}@osu.cz

1 Introduction

There are two standard ways how to understand the concept of fuzzy real numbers. The first one goes back to Zadeh's pioneering paper [9] in 1965 and it is still widely used, see [2] e.g. In this concept fuzzy real numbers are represented as special functions from the set $\mathbb{R}$ of all real numbers to the interval $[0,1]$ (see Definition 2 below) and crisp real numbers are identified with characteristic functions of corresponding singletons. Ten years later Hutton in [1] introduced another way how to understand fuzzy real line, generalizing the representation of real numbers by Dedekind's cuts. This idea was further developed in many papers and the algebraic and topological structures of the fuzzy real line were investigated by several authors, among others by Lowen (e.g. [3], [4]), Rodabaugh ([5], [6]) and Wang (e.g. [7], [8]). In the present paper we define the concept of limit point of a sequence of fuzzy real numbers in the Zadeh's sense and we prove some properties similar to that of limit points of sequences of crisp real numbers.

As usual, in the whole paper we will denote by $\mathbb{N}$ and $\mathbb{R}$ the sets of all positive integers and all real numbers, respectively.

Definition 1. *Let $f\colon \mathbb{R} \to [0,1]$. The kernel $ker(f)$ of f is given by*

$$ker(f) = \{x \in \mathbb{R} \mid f(x) = 1\}.$$

The support $supp(f)$ of f is given by

$$supp\,(f) = \{x \in \mathbb{R} \mid f(x) > 0\}.$$

For each $\alpha \in [0,1]$ the α-cut $[f]_\alpha$ of f is given by

$$[f]_\alpha = \{x \in \mathbb{R} \mid f(x) \geq \alpha\}.$$

Definition 2. *By fuzzy real number we mean any function $f\colon \mathbb{R} \to [0,1]$ with bounded kernel and such that for each $\alpha \in [0,1]$ the α-cut $[f]_\alpha$ is a nonempty convex subset of $\mathbb{R}$.*

Notice that every fuzzy real number, as piecewise monotone function, has at most countably many points of discontinuity. The set of all continuity points of f we denote $\mathcal{C}(f)$.

If f is a fuzzy real number and $x \in \mathbb{R}$ we denote $f(x-)$ and $f(x+)$ the left and right limit of f at x, respectively. Notice that both limits exist at each point. For a fuzzy real number f we denote by $G(f)$ the graph of f. For $(x,y) \in \mathbb{R}^2$ and $\varepsilon > 0$ denote $B_\varepsilon(x,y) = \{(u,v) \in \mathbb{R}^2 \mid \sqrt{(u-x)^2+(v-y)^2} < \varepsilon\}$.

Definition 3. *A sequence (f_n) of fuzzy numbers is bounded if there is an interval $(a,b) \subset \mathbb{R}$ such that $supp\,(f_n) \subset (a,b)$ for every $n \in \mathbb{N}$.*

[1] Supported by the grant MSM6198898701

Definition 4. *A fuzzy number f is said to be a limit point of the sequence (f_n) of fuzzy numbers if the set $I(x,\varepsilon) = \{n \in \mathbb{N} \mid B_\varepsilon(x, f(x)) \cap G(f_n) \neq \emptyset\}$ is infinite for every $\varepsilon > 0$ and $x \in \mathbb{R}$. Denote $L(f_n)$ the set of all limit points of (f_n). A fuzzy number f is said to be a limit of the sequence (f_n) of fuzzy numbers if $L(f_n) = \{f\}$. In this case we write $f_n \rightsquigarrow f$.*

It is easy to see that $L(f_n)$ contains all pointwise limits of subsequences of (f_n). The opposite is not true in general.

Let $\mathcal{F} = (f_n)$ be a sequence of fuzzy real numbers. Denote $\mathcal{A}(\mathcal{F})$ the set of all pairs $(x,y) \in \mathbb{R}^2$ such that for every $\varepsilon > 0$ the set $I = \{n \in \mathbb{N} \mid B_\varepsilon(x,y) \cap G(f_n) \neq \emptyset\}$ is infinite. Notice that a fuzzy real number f is a limit point of a sequence $\mathcal{F} = (f_n)$ if and only if $G(f) \subset \mathcal{A}(\mathcal{F})$.

Lemma 1. *The set $\mathcal{A}(\mathcal{F})$ is nonempty and closed for every sequence of fuzzy real numbers $\mathcal{F} = (f_n)$. All sets of the form $\mathcal{A}(\mathcal{F}) \cap (\{x\} \times [0,1])$, $x \in \mathbb{R}$ are nonempty and closed, the set $\mathcal{A}(\mathcal{F}) \cap (\mathbb{R} \times \{1\})$ is closed and, if $\mathcal{F}$ is bounded, then it is also nonempty.*

2 Results

The following theorem is in accordance with the crisp case.

Theorem 1. *Every bounded sequence of fuzzy numbers has a limit point.*

Although all bounded sequences of fuzzy real numbers have some limit point, the following example shows that it is very unusual for a sequence of fuzzy real numbers to have a limit.

Example 1. Let $f_n = \chi_{(0,1)}$, $n = 1, 2, \ldots$, where χ_A means the characteristic function of the set A. Then $L(f_n) = \{\chi_{(0,1)}, \chi_{(0,1]}, \chi_{[0,1)}, \chi_{[0,1]}\}$. Thus a constant sequence need not have a limit! This suggests the idea to identify all four above functions, which differ each from the other only in points of their discontinuity, to obtain the limit.

The following definition is similar to that in [8], although we use it for rather different class of functions.

Definition 5. *For two fuzzy real numbers f, g define $f \sim g$ if and only if $C(f) = C(g)$ and*

$$f(x-) = g(x-) \qquad \text{and} \qquad f(x+) = g(x+)$$

hold for every $x \in \mathbb{R}$.

Remark 1. It is easy to see that $\sim$ is an equivalence relation on set of all fuzzy numbers. For a fuzzy number f denote the corresponding equivalence class by $[f]$ and denote by $\mathcal{F}(\mathbb{R})$ the set of all equivalence classes of fuzzy real numbers. Notice that $[f]$ is singleton if and only if f is either continuous or of the form $\chi_{\{x\}}$ for some $x \in \mathbb{R}$. Also notice that the condition that f and g have the same set of continuity points, although missing in the corresponding definition in [8], is substantial in Definition 5. Otherwise any two different real numbers $x \neq y$ would be represented by equivalent fuzzy real numbers $\chi_{\{x\}}$ and $\chi_{\{y\}}$ and, consequently, all real numbers would be represented by the same class of equivalence. Finally, notice that the concept of limit point is compatible with the above defined equivalence relation in the sense that if $f_n \sim g_n$ for all $n \in \mathbb{N}$ then $\{[f] \mid f \in L(f_n)\} = \{[g] \mid g \in L(g_n)\}$. Thus Theorem 1 remains valid also in $\mathcal{F}(\mathbb{R})$ and notation $L([f_n])) = \{[f] \mid f \in L(f_n)\}$ makes a sense.

In the structure of real numbers it holds that if x_0 is a limit point of a sequence (x_n) of real numbers then there exists a subsequence (x_{n_k}) of (x_n) such that $x_0 = \lim_{k\to\infty} x_{n_k}$. The following example shows that an analogous statement in the structure of fuzzy real numbers does not hold.

Example 2. Let (q_n) be a sequence of all rational numbers in interval $(0,1)$. For $n \in \mathbb{N}$ let $T_n\colon [0,1] \to [0,1]$ be the function whose graph consists of two line segments connecting pairs of points $(0,0),(q_n,1)$ and $(q_n,1),(1,0)$. Let $f_n\colon (-\infty,\infty) \to [0,1]$ be the extension of T_n defining $f_n(x) = 0$ for all $x \in (-\infty,0) \cup (1,\infty)$ and for every $n \in \mathbb{N}$. Then $[\chi_{(0,1)}] \in L([f_n])$ although there is no subsequence (n_k) with $f_{n_k} \rightsquigarrow \chi_{(0,1)}$.

Now we are interested in such objects as $\liminf$ and $\limsup$ of a given sequence of fuzzy real numbers. It is not satisfactory to order fuzzy real numbers by standard partial order of functions, i.e. $f \le g$ if and only if $f(x) \le g(x)$ holds for all $x \in \mathbb{R}$. One reason is that, if doing so, the natural embedding of $\mathbb{R}$ into $\mathcal{F}(\mathbb{R})$ via the correspondence $x \hookrightarrow \chi_{\{x\}}$ was not order preserving. To keep it order preserving, we define a partial order on the set $\mathcal{F}(\mathbb{R})$ as follows. For every $[f] \in \mathcal{F}(\mathbb{R})$ denote

$$D_-([f]) = \{x \in \mathbb{R} \mid \forall y \in \overline{Ker(f)}\ x < y\} \quad \text{and} \quad D^+([f]) = \{x \in \mathbb{R} \mid \forall y \in \overline{Ker(f)}\ x > y\},$$

where $\overline{A}$ means the closure of a set A. Notice that both $D_-([f])$ and $D^+([f])$ do not depend on the choice of $f \in [f]$, thus we will use also notations $D_-(f)$ and $D^+(f)$ in the sequel.

Definition 6. *For every $[f],[g] \in \mathcal{F}(\mathbb{R})$ define $[f] \preceq [g]$ if and only if*

$$f(x-) \ge g(x-) \quad \text{and} \quad f(x+) \ge g(x+) \quad \text{if} \quad x \in D_-(f)$$

and

$$f(x-) \le g(x-) \quad \text{and} \quad f(x+) \le g(x+) \quad \text{if} \quad x \in D^+(f).$$

In this case we also write $[g] \succeq [f]$.

Remark 2. It is a routine job to check that the relation $\preceq$ does not depend on the choice of $f \in [f]$ and $g \in [g]$ and it is a partial order on $\mathcal{F}(\mathbb{R})$. Notice also that $[f] \preceq [g]$ implies $D_-(f) \subset D_-(g)$ and $D^+(f) \supset D^+(g)$.

As usual, we define a $\liminf$ ($\limsup$) of a sequence $([f_n])$ from $\mathcal{F}(\mathbb{R})$ as the least (greatest) element of the set of all limits points of $([f_n])$. The following theorems are again in accordance with the crisp case.

Theorem 2. *For every bounded sequence $([f_n])$ in $\mathcal{F}(\mathbb{R})$ both $\liminf[f_n]$ and $\limsup[f_n]$ exist.*

Theorem 3. *Every bounded monotone sequence $([f_n])$ in $\mathcal{F}(\mathbb{R})$ has a limit.*

Remark 3. Notice that if $f \sim g$ then for every $\alpha \in [0,1]$ we have

$$\inf[f]_\alpha = \inf[g]_\alpha \qquad \text{and} \qquad \sup[f]_\alpha = \sup[g]_\alpha,$$

where $\inf[f]_\alpha$ and $\sup[f]_\alpha$ mean infimum and supremum of α-cut of f, respectively.

Definition 7. *Let $\mathcal{F} = ([f_n])$ be a sequence in $\mathcal{F}(\mathbb{R})$ with $\underline{[f]} = \liminf[f_n]$ and $\overline{[f]} = \limsup[f_n]$. For $y \in [0,1]$ denote*

$$\psi^+(y) = |\sup\overline{[f]}_y - \sup\underline{[f]}_y| \quad \text{and} \quad \psi^-(y) = |\inf\overline{[f]}_y - \inf\underline{[f]}_y|.$$

If both $\psi^-(y) < \infty$ and $\psi^+(y) < \infty$ for all $y \in (0,1)$, define the dispersion of $\mathcal{F}$ by

$$\delta(\mathcal{F}) = \frac{1}{2}\int_0^1 (\psi^+(y) + \psi^-(y))dy$$

if the integral on the left side is convergent, and define $\delta(\mathcal{F}) = \infty$, otherwise.

Notice that the definition is correct as $\delta(\mathcal{F})$ does not depend on the choice of representatives of $[f_n], [\underline{f}], [\overline{f}]$.

Theorem 4. *The following conditions are equivalent for every sequence $\mathcal{F} = ([f_n])$ in $\mathcal{F}(\mathbb{R})$.*

(i) $\mathcal{F}$ *has a limit,*
(ii) $\limsup[f_n] = \liminf[f_n]$,
(iii) $\delta(\mathcal{F}) = 0$.

References

1. Hutton, B.: Normality in fuzzy topological spaces, J.Math.Anal.Appl. 50, (1975), 74–79.
2. Klement E. P., Mesiar R. and Pap E.: Triangular norms, Trends in Logic–Studia Logica Library 8, Kluwer Academic Publishers, (2000), Dordrecht.
3. Lowen R.: On (**R**(**L**), $\bigoplus$), Fuzzy Sets and Systems 10, (1983), 203–209.
4. Lowen R.: Fuzzy integers, fuzzy rationals and other subspaces of the fuzzy real line , Fuzzy Sets and Systems 14, (1984), 231–236.
5. Rodabaugh S. E.: Fuzzy addition in the L-fuzzy real line, Fuzzy Sets and Systems 8 (1982), 39–52.
6. Rodabaugh S. E.: The L-fuzzy real line and its subspaces, in: R.R Yager (Ed.) Recent Development in Fuzzy Set and Possibility Theory, Pergamon Press Oxford (1982), 402 –418.
7. Wang Geping: Fuzzy addition in the Lowen fuzzy real line, Fuzzy Sets and Systems 27 (1988), 303–315.
8. Wang Geping, Xiaoyong Xi: Convergence of sequences on the fuzzy real line, Fuzzy Sets and Systems 127, (2002), 323–331.
9. Zadeh L. A.: Fuzzy sets, Inform. Control 8, (1965) 338–353.

Convergence of Radon Probability Measures

Ulrich Höhle

FB C Mathematik und Naturwissenschaften
Bergische Universität
D-42097 Wuppertal, Germany
Ulrich.Hoehle@math.uni-wuppertal.de

Let X be a Hausdorff space, $\mathcal{M}_+^1$ be the set of all Radon probability measures on X, and let $\mathcal{G}(X)$ be the set of all lower semicontinuous maps $X \xrightarrow{g} [0,1]$. In 1970, Flemming Topsœ introduced the concept of *weak convergence* of Radon probability measures as follows (cf. [3,4]): A filter $\mathbb{F}$ on $\mathcal{M}_+^1$ converges weakly to a Radon probability measure μ_0 iff for all $g \in \mathcal{G}(X)$ the following inequality holds:

$$\mu_0(g) \;\leq\; \sup_{F \in \mathbb{F}} \Big(\inf_{\mu \in F} \mu(g) \Big) \tag{1}$$

where $\mu(g)$ denotes the Lebesgue integral of g w.r.t. μ. Since the real unit interval $[0,1]$ is *completely distributive*, it is easily seen that this convergence notion is topological — this means that weak convergence coincides with topological convergence w.r.t. the coarsest topology on $\mathcal{M}_+^1$ making all maps $\mu \rightsquigarrow \mu(g)$ lower semicontinuous ($g \in \mathcal{G}(X)$).

Finally, let $X \overset{\delta}{\hookrightarrow} \mathcal{M}_+^1$ be the embedding of X into $\mathcal{M}_+^1$ sending every point $x \in X$ to its Dirac measure δ_x at x. Then the following result is well known.

Theorem. (cf. pp. 371 in [3] and Theorem 11.1 in [4])

1. *$\mathcal{M}_+^1$ is Hausdorff separated in the sense of the topology of weak convergence.*
2. *δ is a homeomorphism onto its range w.r.t. the topology of weak convergence.*
3. *$\delta(X)$ is a closed subset of $\mathcal{M}_+^1$.*
4. *The convex hull of $\delta(X)$ is dense in $\mathcal{M}_+^1$ w.r.t. the topology of weak convergence.*

As an immediate corollary from the previous theorem we obtain the statement that in general the image of X under δ is not dense in $\mathcal{M}_+^1$ w.r.t to the topology of weak convergence. Hence we cannot extend operations on X by the principle of continuous extension to the space $\mathcal{M}_+^1$ of all Radon probability measures on X.

The aim of this talk is to solve this problem and to introduce a *many valued topology* τ on $\mathcal{M}_+^1$ satisfying the following conditions (cf. [1]):

- τ characterizes weak convergence of measures — i.e. a filter $\mathbb{F}$ on $\mathcal{M}_+^1$ converges to a Radon probability measure μ_0 in the sense of τ iff $\mathbb{F}$ converges weakly to μ_0 (i.e. in the sense of (1)).
- $(\mathcal{M}_+^1, \tau)$ is Hausdorff separated.
- $\delta(X)$ is a dense subset of $\mathcal{M}_+^1$ w.r.t. τ.

In this context, a simple example constitutes the real unit interval viewed as the set of all *probability measures* on $2 = \{0,1\}$. In particular, Łukasiewicz negation appears as the unique continuous extension of the classical negation on 2 w.r.t. τ. Further, the formation of image measures w.r.t. continuous maps φ coincides with the unique continuous extension of φ. In this sense many valued topology broadens the space of topological applications to traditional problems of probability theory.

References

1. U. Höhle, *Many Valued Topology and Its Applications* (Kluwer, Boston 2001).
2. U. Höhle, *Many valued topologies and Borel probability measures*, in: S.E. Rodabaugh and E.P. Klement (eds.), Topological and Algebraic Structures in Fuzzy Sets, Trends in Logic **20**, 115-135 (Kluwer Academic Publishers, Dordrecht, Boston, London 2003).
3. L. Schwartz, *Radon Measures* (Tata Institute of Fundamental Research, Oxford University Press 1973).
4. F. Topsœ, *Topology and Measure*, Lecture Notes in Mathematics **133** (Springer-Verlag, Berlin, Heidelberg, New York 1970).

Integrals Which Can Be Defined on Arbitrary Measurable Spaces

Erich Peter Klement[1], Radko Mesiar[2,3], and Endre Pap[4]

[1] Department of Knowledge-Based Mathematical Systems
Johannes Kepler University, 4040 Linz, Austria
ep.klement@jku.at
[2] Department of Mathematics and Descriptive Geometry
Faculty of Civil Engineering, Slovak Technical University
81 368 Bratislava, Slovakia
[3] Institute for Research and Applications of Fuzzy Modeling
University of Ostrava, 701 03 Ostrava, Czech Republic
mesiar@math.sk
[4] Department of Mathematics and Informatics, University of Novi Sad
21000 Novi Sad, Serbia
pap@im.ns.ac.yu, pape@eunet.yu

1 Introduction

We try to contribute to a classical discussion: "What is an integral?" Based on certain minimal sets of axioms we introduce three concepts of functionals (called integral functionals, general and universal integrals), which can be defined on arbitrary measurable spaces and which act on measures which are only (finite) monotone set functions and for measurable functions whose range is contained in the unit interval. Several special types of such functionals, including extremal ones, are characterized.

2 General integral

Throughout of this paper, let X be a fixed non-empty set, $\mathcal{A}$ a σ-algebra of subsets of X (in the case of a finite set X we usually take $\mathcal{A} = 2^X$), and $\mathcal{F}$ the class of all measurable functions $f\colon X \to [0,1]$. Finally, denote by $\mathcal{M}$ the class of all monotone set functions $m\colon \mathcal{A} \to [0,1]$ (considered, sometimes with additional properties, in [4, 6, 11, 14, 15]) which satisfy $m(\emptyset) = 0$, $m(X) = 1$ and $m(A) \le m(B)$ whenever $A \subseteq B$.

Definition 21 A function $\mathbf{I}\colon \mathcal{M} \times \mathcal{F} \to [0,1]$ is called an *integral functional* if it satisfies the following conditions:

(I1) *boundary conditions*, i.e., for each $m \in \mathcal{M}$ we have

$$\mathbf{I}(m,0) = 0 \quad \text{and} \quad \mathbf{I}(m,1) = 1,$$

(I2) *monotonicity* in both coordinates, i.e., for $m_1 \le m_2$ and $f_1 \le f_2$ we have

$$\mathbf{I}(m_1, f_1) \le \mathbf{I}(m_2, f_2),$$

(I3) *extension of the measure*, i.e., for each $A \in \mathcal{A}$ and for each $m \in \mathcal{M}$ we have

$$\mathbf{I}(m, \mathbf{1}_A) = m(A).$$

Obviously, (I3) implies (I1); however we prefer to keep axiom (I1) in order to stress the boundary conditions of integral functionals.

Integral functionals were already discussed in [13], where the extremal integral functionals were given:

$$\mathbf{I}_*(m,f) = \max(\inf f, m(\{f=1\})),$$
$$\mathbf{I}^*(m,f) = \min(\sup f, m(\{f>0\})).$$

Evidently, the class I of all integral functionals is a partially ordered set with smallest element $\mathbf{I}_*$ and greatest element $\mathbf{I}^*$. Moreover, the class I is closed under each idempotent aggregation operator A, i.e., if $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n \in I$, then also $A(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n)$ given by

$$A(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n)(m,f) = A(\mathbf{I}_1(m,f), \mathbf{I}_2(m,f), \ldots, \mathbf{I}_n(m,f))$$

is an element of I. In particular, I is a convex set.

Definition 22 An integral functional $\mathbf{G}\colon \mathcal{M} \times \mathcal{F} \to [0,1]$ is called a *general integral* if it satisfies the following additional conditions:

(I4) *idempotency*, i.e., for every $c \in [0,1]$ we have

$$\mathbf{G}(m,c) = c,$$

(I5) *existence of a pseudo-multiplication*, i.e., there exists a binary operation $\otimes\colon [0,1]^2 \to [0,1]$ such that for each $m \in \mathcal{M}$, each $c \in [0,1]$ and each $A \in \mathcal{A}$

$$\mathbf{G}(m, c \cdot \mathbf{1}_A) = c \otimes m(A),$$

(I6) for each measurable function $\varphi\colon X \to X$, and for each $(m,f) \in \mathcal{M} \times \mathcal{F}$ we have

$$\mathbf{G}(m^\varphi, f) = \mathbf{G}(m, f \circ \varphi),$$

where the measure $m^\varphi \in \mathcal{M}$ is given by $m^\varphi(A) = m(\varphi^{-1}(A))$.

Observe that one of the consequences of axiom (I6) is that for a general integral $\mathbf{G}$ and a Dirac measure $m_{\{x_0\}}$ (all the mass is concentrated in some point $x_0 \in X$) we have $\mathbf{G}(m_{\{x_0\}}, f) = f(x_0)$ for each $f \in \mathcal{F}$.

Clearly, the Choquet, and the Sugeno integral are well-known examples of general integrals, whereas the Lebesgue integral is defined for σ-additive measures $m \in \mathcal{M}$ only.

Also, the class $\mathcal{G}$ of general integrals is convex, and it is closed under each idempotent aggregation operator.

Observe that the axioms (I1)–(I4) imply that the pseudo-multiplication $\otimes$ required in axiom (I5) is a semicopula [1, 5]:

Lemma 23 *Let* $\mathbf{G}$ *be a general integral, and let* $\otimes$ *be the pseudo-multiplication related to* $\mathbf{G}$ *required by* (I5)*. Then* $\otimes$ *is a semicopula, i.e., a binary aggregation operator with neutral element* 1.

Proof: Let $\otimes$ be a pseudo-multiplication related to $\mathbf{G}$ which exists because of (I5). Then (I2) implies the monotonicity of $\otimes$, and from (I1) we derive that 0 is an annihilator of $\otimes$ and $1 \otimes 1 = 1$. Moreover, due to (I3) we have $1 \otimes u = u$ for each $u \in [0,1]$ (observe that, whenever $\operatorname{card}(X) > 1$, for each $A \in \mathcal{A}$ with $\emptyset \subset A \subset X$ we can find an $m_u \in \mathcal{M}$ such that $m_u(A) = u$). Similarly, (I4) implies $c \otimes 1 = c$ for all $c \in [0,1]$. □

Following the ideas of inner and outer measures in classical measure theory, we obtain the following result:

Theorem 24 *Let $\otimes$ be a semicopula. Then the class $\mathcal{G}_\otimes$ of all general integrals related to $\otimes$ is a convex class with smallest element $\mathbf{G}_\otimes$ and greatest element $\mathbf{G}^\otimes$, given by*

$$\mathbf{G}_\otimes(m,f) = \sup\{t \otimes m(\{f \geq t\}) \mid t \in [0,1]\},$$
$$\mathbf{G}^\otimes(m,f) = (\sup f) \otimes m(\{f > 0\}).$$

Proof: It is a matter of direct checking that $\mathbf{G}_\otimes, \mathbf{G}^\otimes \in \mathcal{G}_\otimes$. Moreover, the inequality $\mathbf{G}_\otimes \leq \mathbf{G} \leq \mathbf{G}^\otimes$ for all $\mathbf{G} \in \mathcal{G}_\otimes$ is an immediate consequence of (I2) and (I5). □

Recall that the drastic product $T_\mathbf{D}$ is the weakest and the minimum $T_\mathbf{M}$ is the strongest semicopula. Obviously, for any two semicopulas $\otimes_1$ and $\otimes_2$ with $\otimes_1 \leq \otimes_2$ we have $\mathbf{G}_{\otimes_1} \leq \mathbf{G}_{\otimes_2}$ and $\mathbf{G}^{\otimes_1} \leq \mathbf{G}^{\otimes_2}$.

Corollary 25 *The smallest general integral $\mathbf{G}_* = \mathbf{G}_{T_\mathbf{D}}$ and the largest general integral $\mathbf{G}^* = \mathbf{G}_{T_\mathbf{M}}$ are given by*

$$\begin{aligned}\mathbf{G}_*(m,f) &= \sup\{T_\mathbf{D}(t, m(\{f \geq t\})) \mid t \in [0,1]\}\\ &= \max(\mathrm{essinf}_m f, m(\{f = 1\})),\\ \mathbf{G}^*(m,f) &= \min(\sup f, m(\{f > 0\})).\end{aligned}$$

Note that we have $\mathbf{G}^* = \mathbf{I}^*$ and $\mathbf{G}_* \geq \mathbf{I}_*$, and the inequality in the latter case may be strict.

3 Universal integrals

For $(m,f) \in \mathcal{M} \times \mathcal{F}$, define the function $h_{m,f}\colon [0,1] \to [0,1]$ by

$$h_{m,f}(t) = m(\{x \mid f(x) \geq t\}).$$

Obviously, $h_{m,f}$ is non-increasing (and, therefore, Borel measurable) and satisfies $h_{m,f}(0) = 1$. Following the ideas of [13] (where the name regular integral was used), we introduce another type of integral.

Definition 31 A general integral $\mathbf{U}\colon \mathcal{M} \times \mathcal{F} \to [0,1]$ is called *universal integral* whenever there exists a monotone functional $\mathbf{J}\colon \mathcal{L}([0,1]) \to [0,1]$ such that

$$\mathbf{U}(m,f) = \mathbf{J}(h_{m,f}),$$

where $\mathcal{L}([0,1])$ is the class of all Borel measurable functions from $[0,1]$ to $[0,1]$.

Note that there are two other equivalent concepts of defining universal integrals [7, 13].

Again, the Choquet, and the Sugeno integral are examples of universal integrals. The class $\mathcal{U}$ of universal integrals is also a convex set, and it is closed under any idempotent aggregation operator.

Now we recall two properties of the class $\mathcal{U}$ of universal integrals given in [13]:

(i) For each $\mathbf{U} \in \mathcal{U}$ we have

$$\mathbf{U}_* \leq \mathbf{U} \leq \mathbf{U}^*,$$

where $\mathbf{U}_*$ and $\mathbf{U}^*$ are given by

$$\mathbf{U}_* = \mathbf{G}_*,$$
$$\mathbf{U}^*(m,f) = \min(\mathrm{essup}_m f, m(\{f > 0\})).$$

(ii) For each measurable space $(X, 2^X)$, each $\{0,1\}$-valued measure $m \in \mathcal{M}$ and each $f \in \mathcal{F}$ we have $\mathbf{U}_*(m,f) = \mathbf{U}(m,f) = \mathbf{U}^*(m,f)$ for all $\mathbf{U} \in \mathcal{U}$. Defining the function $L_m \colon \mathcal{F} \to [0,1]$ by $L_m(f) = \mathbf{U}_*(m,f)$, then L_m is a lattice polynomial on X, and it can be written as

$$L_m(f) = \bigvee_{m(A)=1} \bigwedge_{x \in A} f(x).$$

Similarly as in the case of general integrals, we get the following result:

Theorem 32 *Let $\otimes$ be a semicopula. Then the class $\mathcal{U}_\otimes$ of all universal integrals related to $\otimes$ is a convex class with smallest element $\mathbf{U}_\otimes$ and greatest element $\mathbf{U}^\otimes$, given by*

$$\mathbf{U}_\otimes = \mathbf{G}_\otimes,$$
$$\mathbf{U}^\otimes(m,f) = (\mathrm{essup}_m f) \otimes m(\{f > 0\}).$$

Proof: As a consequence of Theorem 24, $\mathbf{G}_\otimes$ satisfies the axioms (I1)–(I6). Moreover, we have for each $(m,f) \in \mathcal{M} \times \mathcal{F}$

$$\mathbf{G}_\otimes(m,f) = \mu\big(\big\{(x,y) \in]0,1[^2 \;\big|\; 0 < y < m(\{f \geq x\})\big\}\big),$$

where the monotone set function $\mu \colon \mathcal{B}(]0,1[^2) \to [0,1]$ is given by

$$\mu(E) = \sup\{t \otimes u \mid]0,t[\times]0,u[\subset E\}.$$

Now it suffices to define $\mathbf{J}_\otimes \colon \mathcal{L}([0,1]) \to [0,1]$ by

$$\mathbf{J}_\otimes(g) = \mu\big(\big\{(x,y) \in]0,1[^2 \;\big|\; 0 < y < g(x)\big\}\big)$$

and to put $\mathbf{U}_\otimes(m,f) = \mathbf{J}_\otimes(h_{m,f})$ (see [7]).

Concerning $\mathbf{U}^\otimes$, it is a matter of straightforward checking only to verify that it satisfies (I1)–(I6). Moreover, $\mathbf{U}^\otimes(m,f) = \mathbf{J}^\otimes(h_{m,f})$, where $\mathbf{J}^\otimes \colon \mathcal{L}([0,1]) \to [0,1]$ is given by $\mathbf{J}^\otimes(g) = (\mathrm{essup}_\lambda g) \cdot \sup\{g > 0\}$, i.e., $\mathbf{U}^\otimes \in \mathcal{U}$. For an arbitrary $(m,f) \in \mathcal{M} \times \mathcal{F}$ define $f^* = (\mathrm{essup}_m f) \cdot \mathbf{1}_{\mathrm{supp} f}$. Evidently, $\mathbf{U}^\otimes(m,f) = \mathbf{U}^\otimes(m,f^*)$. Also, for all $t \in [0,1]$ we have $m(\{f \geq t) \leq m(\{f^* \geq t)$ and, thus,

$$\mathbf{U}(m,f) \leq \mathbf{U}(m,f^*) = \mathbf{U}^\otimes(m,f^*) = \mathbf{U}^\otimes(m,f)$$

for all $\mathbf{U} \in \mathcal{U}$. □

Clearly, for the smallest element $\mathbf{U}_*$ and the largest element $\mathbf{U}^*$ we have $\mathbf{U}_* = \mathbf{U}_{T_\mathbf{D}}$ and $\mathbf{U}^* = \mathbf{U}^{T_\mathbf{M}}$.

The formulas for $\mathbf{U}_\otimes$ and $\mathbf{U}^\otimes$ provide two construction methods for universal integrals based on a given semicopula $\otimes$ as underlying pseudo-multiplication. Note that, e.g., $\mathbf{U}_{T_\mathbf{M}}$ is just the Sugeno integral. For special semicopulas we can give an additional construction method:

Let $C \colon [0,1]^2 \to [0,1]$ be a copula [8, 12], i.e., a semicopula inducing a probability measure μ_C on $\mathcal{B}(]0,1[^2)$ via

$$\mu_C([t_1,t_2[\times [u_1,u_2[) = C(t_2,u_2) - C(t_2,u_1) - C(t_1,u_2) + C(t_1,u_1).$$

Then the functional $\mathbf{U}_{(C)} \colon \mathcal{M} \times \mathcal{F} \to [0,1]$ defined by

$$\mathbf{U}_{(C)}(m,f) = \mu_C\big(\big\{(x,y) \in]0,1[^2 \;\big|\; 0 < y < m(\{f \geq x\})\big\}\big)$$

is a universal integral related to the pseudo-multiplication C. Observe that, for the product copula Π, $\mathbf{U}_{(\Pi)}$ is the Choquet integral, and for the minimum $T_\mathbf{M}$ the universal integral $\mathbf{U}_{(T_\mathbf{M})}$ coincides with the Sugeno integral. For more details about this construction see [7].

Note that, similarly as for logical connectives in fuzzy logics, there is a general method based on automorphisms on $[0,1]$ for constructing new types of integrals from a given one:

Proposition 33 *Let* $\varphi\colon [0,1] \to [0,1]$ *be an increasing bijection and* $\mathbf{I}$ *be an integral functional. Then the functional* $\mathbf{I}_\varphi\colon \mathcal{M} \times \mathcal{F} \to [0,1]$ *defined by*

$$\mathbf{I}_\varphi(m,f) = \varphi^{-1}(\mathbf{I}(\varphi \circ m, \varphi \circ f))$$

is an integral functional. Moreover, $\mathbf{I}_\varphi$ *is a general integral or a universal integral whenever* $\mathbf{I}$ *is a general integral or a universal integral, respectively.*

The proof of this result is obvious and, therefore, omitted. Observe only that, if a general integral $\mathbf{G}$ is related to the semicopula $\otimes$, then $\mathbf{G}_\varphi$ is related to the transformed semicopula $\otimes_\varphi$ given by

$$x \otimes_\varphi y = \varphi^{-1}(\varphi(x) \otimes \varphi(y)).$$

Example 34 The Choquet integral $\mathbf{Ch}$ [3] is a universal integral related to the standard product because of $\mathbf{Ch}(m, c \cdot \mathbf{1}_A) = c \cdot m(A)$. For each $p \in \,]0,\infty[$, define the automorphism $\varphi_p\colon [0,1] \to [0,1]$ by $\varphi_p(x) = x^p$. Then $(\mathbf{Ch}_{\varphi_p})_{p \in]0,\infty[}$ is a parameterized family of universal integrals given by

$$\mathbf{Ch}_{\varphi_p}(m,f) = (\mathbf{Ch}(m^p, f^p))^{\frac{1}{p}}.$$

This family is non-increasing with respect to the parameter, and each integral $\mathbf{Ch}_{\varphi_p}$ is related to the product Π, i.e., $\mathbf{Ch}_{\varphi_p} \in \mathcal{U}_\Pi$. For the limit cases we get

$$\mathbf{Ch}_{\varphi_\infty} = \lim_{p \to \infty} \mathbf{Ch}_{\varphi_p} = \mathbf{U}_\Pi,$$
$$\mathbf{Ch}_{\varphi_0} = \lim_{p \to 0^+} \mathbf{Ch}_{\varphi_p} = \mathbf{U}^\Pi,$$

which means that $\mathbf{Ch}_{\varphi_\infty}$ and $\mathbf{Ch}_{\varphi_0}$ are the smallest and largest universal integral related to Π, respectively. Observe that $\mathbf{Ch}_{\varphi_\infty}$ is the Shilkret integral [10].

There are other approaches for the construction of universal integrals (e.g. the one presented in [2]).

4 Conclusion

We have proposed and studied three general frameworks for integrals which can be defined on arbitrary measure spaces, including some construction methods and the characterization of some extremal integrals. For instance, the greatest and smallest general and universal integrals related to some semicopula are identified.

5 Acknowledgment

The second and the third author were supported by a technology transfer project of the Upper Austrian Government. The second author also acknowledges the support of the Österreichische Forschungsgemeinschaft and by the grants VEGA 1/4209/07, APVT-20-003204, and MSM VZ 6198898701, the third author was also supported by Vojvodina PSSTD and by grant MNTRS-144012.

References

1. B. Bassan and F. Spizzichino, *Relations among univariate aging, bivariate aging and dependence for exchangeable lifetimes*, J. Multivariate Anal. **93** (2005), 313–339.
2. P. Benvenuti, R. Mesiar, and D. Vivona, *Monotone set functions-based integrals*, in Pap [9], pp. 1329–1379.
3. G. Choquet, *Theory of capacities*, Ann. Inst. Fourier (Grenoble) **5** (1953–1954), 131–292.
4. D. Denneberg, *Non-additive measure and integral*, Kluwer Academic Publishers, Dordrecht, 1994.
5. F. Durante and C. Sempi, *Semicopulæ*, Kybernetika (Prague) **41** (2005), 315–328.
6. G. H. Greco, *Monotone integrals*, Rend. Sem. Mat. Univ. Padova **57** (1977), 149–166.
7. E. P. Klement, R. Mesiar, and E. Pap, *Measure-based aggregation operators*, Fuzzy Sets and Systems **142** (2004), 3–14.
8. R. B. Nelsen, *An introduction to copulas*, Lecture Notes in Statistics, vol. 139, Springer, New York, 1999.
9. E. Pap (ed.), *Handbook of measure theory*, Elsevier Science, Amsterdam, 2002.
10. N. Shilkret, *Maxitive measure and integration*, Indag. Math. **33** (1971), 109–116.
11. J. Šipoš, *Integral with respect to a pre-measure*, Math. Slovaca **29** (1979), 141–155.
12. A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8** (1959), 229–231.
13. P. Struk, *Extremal fuzzy integrals*, Soft Computing **10** (2006), 502–505.
14. M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. thesis, Tokyo Institute of Technology, 1974.
15. Z. Wang and G. J. Klir, *Fuzzy measure theory*, Plenum Press, New York, 1992.

Label Semantics as a Framework for Linguistic Models in Data Mining

Jonathan Lawry

Department of Engineering Mathematics
University of Bristol, Bristol, BS8 1TR, UK
j.lawry@bris.ac.uk

1 Introduction

Technological advances and rapidly improving communications have brought about a 'data explosion' in science, engineering, business and finance. There is currently an almost continual generation of large, multi-dimensional databases describing complex systems. However, data alone is useless without methods for extracting important relationships between attributes and parameters that can help us understand and predict the behaviour of the system. For this we need algorithms that can automatically learn models from databases which are sufficiently transparent so that they can be readily understood by domain experts while maintaining good predictive capabilities.

The label semantics framework gives a probabilistic interpretation of the uncertainty resulting from the use of vague or imprecise linguistic descriptions in high-level models. As such it is straightforward to integrate with other sources of uncertainty in data, to provide a representation framework for new learning algorithms which generate linguistic models from data mining. These relatively transparent models allow for a qualitative understanding of the patterns and relationships underlying the data, in addition to giving accurate quantitative predictions.

2 Appropriateness Measures and Mass Assignments

Label semantics proposes two fundamental and inter-related measures of the appropriateness of labels as descriptions of an object or value. Given a finite set of labels LA from which can be generated a set of expressions LE through recursive applications of logical connectives, the measure of appropriateness of an expression $\theta \in LE$ as a description of instance x is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective belief that θ can be used to describe x based on his/her (partial) knowledge of the current labelling conventions of the population. From an alternative perspective, when faced with an object to describe, an agent may consider each label in LA and attempt to identify the subset of labels that are appropriate to use. Let this set be denoted by $\mathcal{D}_x$. In the face of their uncertainty regarding labelling conventions the agent will also be uncertain as to the composition of $\mathcal{D}_x$, and in label semantics this is quantified by a probability mass function $m_x : 2^{LA} \to [0,1]$ on subsets of labels. The relationship between these two measures will be described below.

Definition 1. *Label Expressions*
The set of label expressions of LA, LE, is defined recursively as follows:

- *If* $L \in LA$ *then* $L \in LE$
- *If* $\theta, \varphi \in LE$ *then* $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi \in LE$

A mass assignment m_x on sets of labels then quantifies the agent's belief that any particular subset of labels contains all and only the labels with which it is appropriate to describe x.

Definition 2. *Mass Assignment on Labels*
$\forall x \in \Omega$ *a mass assignment on labels is a function* $m_x : 2^{LA} \rightarrow [0,1]$ *such that* $\sum_{S \subseteq LA} m_x(S) = 1$

Now depending on labelling conventions there may be certain combinations of labels which cannot all be appropriate to describe any object. For example, *small* and *large* cannot both be appropriate. This restricts the possible values of $\mathcal{D}_x$ to the following set of focal elements:

Definition 3. *Set of Focal Elements*
Given labels LA together with associated mass assignment $m_x : \forall x \in \Omega$*, the set of focal elements for LA is given by* $\mathcal{F} = \{S \subseteq LA : \exists x \in \Omega,\ m_x(S) > 0\}$

The appropriateness measure, $\mu_\theta(x)$, and the mass m_x are then related to each other on the basis that asserting 'x is θ' provides direct constraints on $\mathcal{D}_x$. In general we can recursively define a mapping $\lambda : LE \rightarrow 2^{2^{LA}}$ from expressions to sets of subsets of labels, such that the assertion 'x is θ' directly implies the constraint $\mathcal{D}_x \in \lambda(\theta)$ and where $\lambda(\theta)$ is dependent on the logical structure of θ.

Definition 4. *λ-mapping*
$\lambda : LE \rightarrow 2^{2^{LA}}$ *is defined recursively as follows:* $\forall \theta,\ \varphi \in LE$

- $\forall L_i \in LA\ \lambda(L_i) = \{T \subseteq LA : L_i \in T\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$, $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$, $\lambda(\neg\theta) = \lambda(\theta)^c$

Based on the λ mapping we then define $\mu_\theta(x)$ as the sum of m_x over those set of labels in $\lambda(\theta)$.

Definition 5. *Appropriateness Measure*

$$\forall \theta \in LE,\ \forall x \in \Omega\ \mu_\theta(x) = \sum_{S \in \lambda(\theta)} m_x(S)$$

Appropriateness measures are not in general functional since m_x cannot be uniquely determined from $\mu_L(x) : L \in LA$. However, in the presence of additional assumptions the calculus can be functional (although never truth-functional). One such assumption, based on an idea of ordering which is often rather natural for labels defined in data models, is as follows:

Definition 6. *Consonance in Label Semantics*
Given non-zero appropriateness measures on basic labels $\mu_{L_i} : i = 1, \ldots, n$ *ordered such that* $\mu_{L_i}(x) \geq \mu_{L_{i+1}}(x)$ *for* $i = 1, \ldots, n$ *then the consonant mass assignment has the form:*

$$m_x(\{L_1, \ldots, L_n\}) = \mu_{L_n}(x),\ m_x(\emptyset) = 1 - \mu_{L_1}(x)$$
$$m_x(\{L_1, \ldots, L_i\}) = \mu_{L_i}(x) - \mu_{L_{i+1}}(x)\ \textit{for}\ i = 1, \ldots, n$$

In this context the consonant assumption is that for each $x \in \Omega$ an agent first identifies a total ordering on the appropriateness of labels. They then evaluate their belief values m_x about which labels are appropriate to describe x in such a way so as to be consistent with this ordering. Given definition 5 and the consonance assumption it can be shown that appropriateness measures have the following general properties [1], [2], [5]:

Theorem 1. *Properties of Appropriateness Measures*
$\forall \theta, \varphi \in LE$ *then the following properties hold:*

- *If* $\theta \models \varphi$ *then* $\forall x \in \Omega\ \mu_\theta(x) \leq \mu_\varphi(x)$
- *If* $\theta \equiv \varphi$ *then* $\forall x \in \Omega\ \mu_\theta(x) = \mu_\varphi(x)$
- *If* θ *is a tautology then* $\forall x \in \Omega\ \mu_\theta(x) = 1$ *and* $\mu_{\neg\theta}(x) = 0$
- $\forall x \in \Omega\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$
- *Let* $LE^{\wedge,\vee} \subseteq LE$ *denote those expressions generated recursively from LA using only the connectives* $\wedge$ *and* $\vee$. $\forall \theta, \varphi \in LE^{\wedge,\vee}$, $\forall x \in \Omega$ *it holds that:* $\mu_{\theta\wedge\varphi}(x) = \min\left(\mu_\theta(x), \mu_\varphi(x)\right)$ *and* $\mu_{\theta\vee\varphi}(x) = \max\left(\mu_\theta(x), \mu_\varphi(x)\right)$.

3 Linguistic Models from Data

Consider the following formalization of a learning problem: Given attributes $x_1, \ldots, x_{k+1}$ with universes $\Omega_1, \ldots, \Omega_{k+1}$ suppose that x_{k+1} is dependent on $x_1, \ldots, x_k$ according to some functional mapping $g : \Omega_1 \times \ldots \times \Omega_k \to \Omega_{k+1}$ (i.e. $x_{k+1} = g(x_1, \ldots, x_k)$). In the case that Ω_{k+1} is finite then this is referred to as a classification problem whereas if Ω_{k+1} is an infinite subset of $\mathbb{R}$ (typically a closed interval) then it is referred to as a prediction or regression problem. Information regarding this function is then provided by a training database containing vectors of input values together with their associated output. Let this database be denoted by $DB = \{\langle x_1(i), \ldots, x_k(i), x_{k+1}(i)\rangle : i = 1, \ldots, N\}$.

Label semantics can be used to infer models from DB which have a linguistic rule based structure and which provide an approximation $\hat{g}$ of the underlying function mapping g. Here we consider two such models; mass relations and linguistic decision trees. For both approaches we use appropriateness measures to define a set of labels describing each attribute $LA_j : j = 1, \ldots, k+1$ with associated label expressions $LE_j : j = 1, \ldots, k+1$ and focal sets $\mathcal{F}_j : j = 1, \ldots, k+1$.

Definition 7. *Mass Relations*
A mass relation is a conditional function $m : 2^{LA_1} \times \ldots \times 2^{LA_k} \to [0,1]$ *such that for* $F_i \in \mathcal{F}_i : i = 1, \ldots, k+1$

$$m(F_1, \ldots, F_k | F_{k+1}) = \frac{\sum_{i \in DB} \prod_{j=1}^{k+1} m_{x_j(i)}(F_j)}{\sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1})}$$

A mass relation generates a set of weighted rules of the form:

$$(\mathcal{D}_{x_1} = F_1) \wedge \ldots \wedge (\mathcal{D}_{x_k} = F_k) \to (\mathcal{D}_{x_{k+1}} = F_{k+1}) : w \text{ where}$$
$$w = m(F_{k+1}|F_1, \ldots, F_k) = \frac{m(F_1, \ldots, F_k|F_{k+1}) m(F_{k+1})}{\sum_{F_{k+1}} m(F_1, \ldots, F_k|F_{k+1}) m(F_{k+1})}$$
$$\text{and } m(F_{k+1}) = \frac{1}{N} \sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1})$$

In practice it can be computationally expensive to calculate the mass relation exactly and typically we need to use some form of approximation. One approach is to search for dependency groupings amongst the attributes and assume conditional independence (given F_{k+1}) between them (see [4] for details).

Definition 8. *Linguistic Decision Trees (LDT)*
A linguistic decision tree is a decision tree where the nodes are attributes from $x_1, \ldots, x_k$ *and the edges are label expressions describing each attribute. More formally, supposing that the j'th node at depth d is the attribute* x_{j_d} *then there is a set of label expressions* $\mathcal{L}_{j,d} \subseteq LE_i$ *forming the edges from* x_{j_d} *such*

that $\lambda(\bigvee_{\theta\in\mathcal{L}_{j,d}}\theta)\supseteq\mathcal{F}_{j_d}$ *and* $\forall\theta,\varphi\in\mathcal{L}_{j,d}\ \lambda(\theta\wedge\varphi)\cap\mathcal{F}_{j_d}=\emptyset$. *Also a branch B from a LDT consists of a sequence of expressions* $\varphi_1,\ldots,\varphi_m$ *where* $\varphi_d\in\mathcal{L}_{j,d}$ *for some* $j\in\mathbb{N}$ *for* $d=1,\ldots,m$, *augmented by a conditional mass value* $m(F_{k+1}|B)$ *for every output focal set* $F_{k+1}\in\mathcal{F}_{k+1}$. *Hence, every branch B encodes a set of weighted linguistic rules of the form:*

$$(x_{j_1}\ is\ \varphi_1)\wedge\ldots\wedge(x_{j_m}\ is\ \varphi_m)\rightarrow(\mathcal{D}_{x_{k+1}}=F_{k+1}):m(F_{k+1}|B)$$

where x_{j_d} *is a depth d attribute node.*

Also the mass assignment value $m(F_{k+1}|B)$ *can be determined from DB according to:*

$$m(F_{k+1}|B)=\frac{\sum_{i\in DB}m_{x_{k+1}(i)}(F_{k+1})\prod_{d=1}^{m}\mu_{\varphi_d}(x_{j_d}(i))}{\sum_{i\in DB}\prod_{d=1}^{m}\mu_{\varphi_d}(x_{j_d}(i))}$$

The LID3 algorithm has been developed as an entropy guided algorithm for learning the structure of an LDT from data (see [3] for details).

References

1. J. Lawry, (2004), 'A Framework for Linguistic Modelling' *Artificial Intelligence*, Vol. 155, pp1-39
2. J. Lawry, (2006), *Modelling and Reasoning with Vague Concepts*, Springer
3. Z. Qin, J. Lawry, (2005), 'Decision Tree Learning with Fuzzy Labels', *Information Sciences*, Vol. 172, pp91-129
4. N. Randon, J. Lawry, (2006), 'Classification and Query Evaluation using Modelling with Words', *Information Sciences*, Vol. 176, pp438-464
5. Y. Tang, J. Zheng, (2006), 'Linguistic Modelling Based on Semantic Similarity Relation Among Linguistic Labels', *Fuzzy Sets and Systems*, Vol. 157, pp1662-1673

Probability Theory on IF-Events With Applications in Other Fuzzy and Quantum Structures

Katarína Lendelová and Beloslav Riečan

Department of Mathematics
Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 97401 Banská Bystrica, Slovakia
{lendelov|riecan}@fpv.umb.sk

The communication consists of two parts. The first one is dedicated to IF-events and it serves as a motivation. Of course, IF-sets seems to be important from the point of view of applications and therefore the construction of the probability theory is very important. Recall that an IF-set is a pair $A = (\mu_A, \nu_A)$ of fuzzy sets such that $\mu_A + \nu_A \leq 1$; μ_A is the membership function, ν_A is the non-membership function.

We present two ways. The first one ([3]) is based on the Lukasiewicz connectives $A \oplus B = (\mu_a \oplus \mu_B, \nu_A \odot \nu_B)$, $A \odot B = (\mu_a \odot \mu_B, \nu_A \oplus \nu_B)$, where $f \oplus g = min(f+g, 1)$, $f \odot g = max(f+g-1, 0)$. The crucial notion is the additivity:

$$P(A) + P(B) = P(A \oplus B) + P(A \odot B).$$

This theory can be embedded to the well extended MV-algebra probability theory ([4]).

Of course, recently another approach was proposed ([1]) characterized by the additivity:

$$P(A) + P(B) = P(A \cup B) + P(A \cap B).$$

In this case the results of MV-algebra theory can not be applied, of course, some methods of this theory can be used.

Also the preceding facts justify a general approach considering the probability as a mapping from an ordered set $(M, \oplus, \odot)$ with arbitrary operations $\oplus$, $\odot$ to the unit interval ([2]). The results of such general theory can be applied to a large variety of known structures, not only IF-events and MV-algebras, but also D-posets, effect algebras and some generalizations of these systems appeared in recent time.

References

1. M. Krachounov (2006). Intuitionistic probability and intuitionistic fuzzy sets. In *First International Workshop on Intuitionistic Fuzzy sets, Generalized Nets and Knowledge Engineering (E. El-darzi, K. Atanassov, P. Chountas, Eds.)*, University of Westminster, pages 18-24, London, UK, 6-7 September 2006.
2. K. Lendelová (2005). Measure theory on multivalued logics and its applications. In *PhD thesis*, FPV UMB, Banská Bystrica, 2005.
3. B. Riečan (2006). On a problem of Radko Mesiar: general form of IF-probabilities. In: *Fuzzy sets and Systems*, volume 152, pages 1485-1900, 2006.
4. B. Riečan - D. Mundici (2002). Probability in MV-algebras. In *Handbook of Measure Theory (E. Pap ed.)*, Elsevier, Amsterdam, pages 869-909, 2002.

On the Extension of Group-Valued Measures

Alžbeta Michalíková

Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 974 01 Banská Bystrica, Slovakia
michalik@fpv.umb.sk

The Carathéodory construction of a measure is well known. Recently it was generalized for the lattice group-valued mappings. In this contribution we consider group-valued case for mapping defined on a space of real functions.

First some properties of outer measures are investigated including the properties of measurable elements. Then the Choquet lemma is proved for the induced outer measure. The results are applied for the extension of a measure defined on a family of functions.

References

1. Carathéodory, C.: Vorelsungen über reele Funktionen. Leipzig, Berlin 1927.
2. Fremlin, D.H.: A direct proof of the Matthes - Wright integral extension theorem. J. London Math. Soc. 11, 1975, 276 - 284.
3. Fremlin, D.H.: Topological Riesz Spaces and Measure Theory. Cambridge Univ. Press 1974.
4. Riečan, B.: Carathéodory measurability revisites. Tatra Moutnains Math. Publ., submitted.
5. Riečan, B.: On measures and integrals with values in ordered groups. Math. Slovaca 33, 1983, 153 -163.
6. Riečan, B. - Neubrunn, T. (1997): Integral, Measure, and Ordering. *Kluwer Academic Publishers, Dordrecht and Ister Science*, Bratislava.
7. Wright. J.D.M.: Stone-algebra-valued measures for vector lattices. Ann. Inst. Fourier (Grenoble) 21, 1971, 65 - 85.

Probability and Conditional Probability on Tribes of Fuzzy Sets

Mirko Navara*

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Praha, Czech Republic
navara@cmp.felk.cvut.cz

Abstract. The classical probability theory and statistics tries to avoid experiments whose results are vague. However, also this type of information is useful and used in everyday life. To treat it, an extension of probability theory to events described by fuzzy sets is highly desirable. An intuitive definition has been suggested by Zadeh [26]. Later on, an axiomatic approach has been introduced by Butnariu and Klement [3]. Independently, measures on MV-algebras have been studied. Particularly interesting results have been obtained by Riečan and Mundici for MV-algebras with product [23]. We show that these two approaches overlap significantly. Further, we outline conditional probabilities in this context.

Keywords: Probability measure, fuzzy set, tribe, MV-algebra, MV-algebra with product, conditional probability.

1 Intuitive definition of probability of fuzzy events

Zadeh [26] suggested to define the probability of a fuzzy set f by the formula

$$\mu(f) = \int f \, dP, \tag{1}$$

where P is a classical probability measure on some σ-algebra $\mathcal{A}$. This is a natural extension of the notion used for crisp sets. However, no justification for exactly this formula was given. To clarify the meaning of the integral, the universe (=the domain of f) must be specified and f has to be measurable w.r.t. the σ-algebra $\mathcal{A}$. This limits the structure of fuzzy sets on which probability can be defined. This will be specified in the subsequent sections. [1]

2 Tribes

An axiomatic approach imitating the classical probability theory was suggested by Höhle [10] and developed by Butnariu and Klement [3]. As an analog of a σ-algebra, they suggest a *tribe* of fuzzy sets. The original notion by Butnariu and Klement [3] is slightly modified in [22] as follows: Let X be a non-empty set. A *tribe* on X is a pentuplet $(\mathcal{T}, \odot, ', 0, \leq)$, where $\mathcal{T} \subseteq [0,1]^X$, $\odot$ is a triangular norm (t-norm for short), $'$ is a strong fuzzy negation, 0 is the constant zero function on X, $\leq$ is the fuzzy inclusion, and the following conditions are satisfied:

(T1) $0 \in \mathcal{T}$,
(T2) $f \in \mathcal{T} \implies f' \in \mathcal{T}$,
(T3) $f, g \in \mathcal{T} \implies f \odot g \in \mathcal{T}$,
(T4) $(f_n)_{n\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$, $f_n \nearrow f \implies f \in \mathcal{T}$

* The author was supported by the project MSM 6840770038 Decision Making and Control in Manufacturing III.

[1] There are many alternative approaches to measure theory on fuzzy sets, see, e.g., [25] and [24], where an extensive bibliography can be found.

(the symbol $\nearrow$ denotes monotone increasing convergence). Butnariu and Klement admitted only the *standard negation* ($\alpha \mapsto 1-\alpha$) for $'$ and, instead of (T3), (T4), they assumed

(T3+) $(f_n)_{n\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}} \implies \bigodot_{n\in\mathbb{N}} f_n \in \mathcal{T}$.

This condition is more general, but the difference is not essential. All results by Butnariu and Klement refer to tribes which satisfy our definition, too.

In the definition of a tribe, we admit an arbitrary strong (=involutive) fuzzy negation. However, in the sequel we restrict attention only to the standard negation. This special case simplifies the formulation of results and it can be easily extended to the general case (see [22] for detailed arguments).

When there is no risk of confusion, we speak briefly of a tribe $(\mathcal{T},\odot)$ (as in [2]), resp. of an $\odot$-*tribe* $\mathcal{T}$ (as in [3]). Our notation follows the pattern of [24]. We also speak of an $\odot$-tribe when we need to refer to the t-norm $\odot$, but not to the tribe itself. In particular cases, when $\odot$ is the *product t-norm*, $x \odot_{\mathbf{P}} y = x\cdot y$, resp. the *Łukasiewicz t-norm*, $x \odot_{\mathbf{L}} y = \max(0, x+y-1)$, we speak of a *product tribe*, resp. a *Łukasiewicz tribe*.

By $\oplus_{\mathbf{P}}$, resp. $\oplus_{\mathbf{L}}$, we denote the triangular conorm (t-conorm for short) dual to $\odot_{\mathbf{P}}$, resp. $\odot_{\mathbf{L}}$ (and similarly for other t-norms and t-conorms). In the sequel, particular attention will be paid to *Frank t-norms* $\odot^{\mathbf{F}}_{\lambda}$, $\lambda \in [0,\infty]$, which are defined by

$$x \odot^{\mathbf{F}}_{\lambda} y = \log_{\lambda}\left(1 + \frac{(\lambda^x - 1)(\lambda^y - 1)}{\lambda - 1}\right)$$

if $\lambda \in \,]0,\infty[\, \setminus \{1\}$, $\odot^{\mathbf{F}}_0 = \odot_{\mathbf{M}} = \min$, $\odot^{\mathbf{F}}_1 = \odot_{\mathbf{P}}$, and $\odot^{\mathbf{F}}_{\infty} = \odot_{\mathbf{L}}$ (see [9, 13] or [12, Example 2.9.3]). They are Archimedean for $\lambda > 0$ and strict for $\lambda \in \,]0,\infty[$.

For a tribe $\mathcal{T}$, we denote by $\mathcal{B}(\mathcal{T})$ its largest Boolean sub-σ-algebra. It consists of all crisp sets of $\mathcal{T}$.

3 Probability on MV-algebras

Another approach to measures of fuzzy sets is based on MV-algebras. (For basic definitions and more details, we refer to [6].) For an MV-algebra $\mathcal{T}$, we denote by $\mathcal{B}(\mathcal{T})$ its largest Boolean subalgebra. It consists of all $b \in \mathcal{T}$ which are *Boolean*, i.e., $b \oplus b = b$. The Boolean algebra $\mathcal{B}(\mathcal{T})$ is also called the *Boolean skeleton* of $\mathcal{T}$. If $\mathcal{T}$ is σ-complete, then $\mathcal{B}(\mathcal{T})$ is a σ-complete Boolean algebra.

A *state*[2] on a σ-complete MV-algebra $\mathcal{T}$ is a mapping $\mu\colon \mathcal{T} \to [0,1]$ satisfying the following conditions:

(S1) $\mu(1) = 1$,
(S2) $f, g \in \mathcal{T},\ f \odot_L g = 0 \implies \mu(f \oplus_L g) = \mu(f) + \mu(g)$,
(S3) $(f_n)_{n\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}},\ f_n \nearrow f \implies \mu(f_n) \to \mu(f)$.

The state space (=the set of all states) is a convex set; its extreme points are called *pure states*.

Theorem 1. *Let $\mathcal{T}$ be a σ-complete MV-algebra. Then every state μ on $\mathcal{T}$ is of the form (1), where $P = \mu \upharpoonright \mathcal{B}(\mathcal{T})$ is an ordinary probability measure on the σ–complete Boolean algebra $\mathcal{B}(\mathcal{T})$.*

This definition and characterization of states applies also to Łukasiewicz tribes, which form a special class of MV-algebras. In this case, the Boolean skeleton coincides with the set of all crisp sets from the tribe, thus the two meanings of $\mathcal{B}(\mathcal{T})$ coincide. Łukasiewicz tribes are characterized in [7]

[2] Usually a state is a synonym for a probability measure. Here we use these notions in different contexts and we distinguish them. However, we shall see that they coincide in important cases.

as those σ-complete MV-algebras which admit a *separating set* of pure states, i.e., for each $a,b \in \mathcal{T}$, $a \neq b$, there is a pure state s such that $s(a) \neq s(b)$. Although not all σ-complete MV-algebras satisfy this property, it is a reasonable requirement in probability theory. This has a Boolean analogy: Among general σ-complete Boolean algebras, only σ-algebras of subsets of a set are usually considered a good basis of a probability theory. The relation of Łukasiewicz tribes to σ-complete MV-algebras is the same as that of σ-algebras to general σ-complete Boolean algebras.

4 Axiomatic approach to probability on tribes

There is a much general way of defining (probability) measures on (not necessarily Łukasiewicz) tribes. Following Butnariu and Klement [3], a *probability measure* on a tribe $(\mathcal{T},\odot,',0,\leq)$ is a functional $\mu\colon \mathcal{T} \to [0,1]$ such that

(M1) $\mu(0)=0, \mu(1)=1$,

(M2) $f,g \in \mathcal{T} \implies \mu(f \odot g)+\mu(f \oplus g)=\mu(f)+\mu(g)$, where $\oplus$ is the t-conorm dual to $\odot$,

(M3) $(f_n)_{n\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}},\ f_n \nearrow f \implies \mu(f_n) \to \mu(f)$.

If, moreover, μ satisfies

(M4) $(f_n)_{n\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}},\ f_n \searrow f \implies \mu(f_n) \to \mu(f)$,

it is called a σ*-order continuous probability measure*.

Theorem 2. *[19] Let $\odot^{\mathbf{F}}_{\lambda}$, $\lambda \in]0,\infty]$, be an Archimedean Frank t-norm. Let $\mathcal{T}$ be an $\odot^{\mathbf{F}}_{\lambda}$-tribe. Every element of $\mathcal{T}$ is $\mathcal{B}(\mathcal{T})$-measurable. Every probability measure on $\mathcal{T}$ is a convex combination of a measure of the form (1) and a measure*

$$\nu(f)=P(\operatorname{supp} f), \qquad f \in \mathcal{T}, \tag{2}$$

where $\operatorname{supp} f=\{x \in X \mid f(x)>0\}$ *is the* support *of a fuzzy set f and $P=\mu \restriction \mathcal{B}(\mathcal{T})$ is an ordinary probability measure on the* σ*-algebra $\mathcal{B}(\mathcal{T})$. Every* σ*-order continuous probability measure on $\mathcal{T}$ is of the form (1).*

A measure of the form (2) is called a *support measure* [3]. Its values depends only on the support, it does not distinguish among positive membership degrees. This seems to be hardly motivated by applications. This is why σ-order continuity was required and condition (M4) has been added in [22]. In [3], (M4) was omitted because in σ-algebras it follows from (M3). However, in tribes this difference is important.

In particular, Th. 2 applies to product or Łukasiewicz tribes. Due to properties of Łukasiewicz operations, the conjunction of (S1), (S2) is equivalent to the conjunction of (M1), (M2) (for strict t-norms, (S2) is much weaker than (M2) and such a condition does not seem useful). Condition (S3) is identical to (M3) and it implies also (M4) in case of a σ-complete MV-algebra. As a consequence, the notions of state and σ-order continuous probability measure coincide for Łukasiewicz tribes. We shall see that this happens in a much more general context.

5 Probability on MV-algebras with products

One important feature of MV-algebras (not achieved in other fuzzifications of Boolean algebras) is the existence of partitions of unity and their joint refinements [6]. However, two partitions of unity need not admit the coarsest joint refinement, and even if it exists, there is no canonical formula for it (see [16] for more details). Therefore the basics of statistics (central limit theorem, etc.) were developed in [23] only for the case when $\mathcal{T}$ is an *MV-algebra with product*, i.e., a pair $(\mathcal{T},\cdot)$, where $\mathcal{T}$ is an MV-algebra and $\cdot$ is a commutative and associative binary operation on $\mathcal{T}$ satisfying:

(P1) $1 \cdot a = a$,

(P2) $a \cdot (b \odot_L c') = (a \cdot b) \odot_L (a \cdot c)'$.

For the standard MV-algebra, $[0,1]$ with the Łukasiewicz operations, the algebraic product is the only operation making it an MV-algebra with product [23]. More generally, a Łukasiewicz tribe forms an MV-algebra with product iff it is equipped with the algebraic product as $\cdot$; then we call it a *Łukasiewicz tribe with product*. The introduction of the product admits to compute joint refinements canonically.

Two above approaches coincide in the following important case:

Theorem 3. *Let $\odot^{\mathbf{F}}_{\lambda}$, $\lambda \in]0,\infty[$, be a strict Frank t-norm. Then $\odot^{\mathbf{F}}_{\lambda}$-tribes are exactly Łukasiewicz tribes with product (when equipped with the respective operations); states on them coincide with σ-order continuous probability measures.*

It is surprising that we obtain the same notion from two axiomatic systems where condition (S2) (formulated for the *Łukasiewicz* t-norm) is much different from condition (M2) (formulated for the *product* t-norm, resp. a *strict Frank* t-norm).

As a corollary, Łukasiewicz tribes with product are exactly product tribes.

There are also characterizations of probability measures on other tribes.

Theorem 4. *[22] Let $\odot$ be a (strict) t-norm of the form*

$$\alpha \odot \beta = h^{-1}(h(\alpha) \odot^{\mathbf{F}}_{\lambda} h(\beta)), \tag{3}$$

where $\odot^{\mathbf{F}}_{\lambda}$, $\lambda \in]0,\infty[$, is a strict Frank t-norm and $h\colon [0,1] \to [0,1]$ is an increasing bijection which commutes with the standard negation, i.e., $h(\alpha') = (h(\alpha))'$ for all α. Then $\odot$-tribes are exactly Łukasiewicz tribes with product. Let $\mathcal{T}$ be an $\odot$-tribe. Every element of $\mathcal{T}$ is $\mathcal{B}(\mathcal{T})$-measurable. Each σ-order continuous probability measure on $\mathcal{T}$ is of the form

$$\mu(f) = \int h \circ f \, dP, \tag{4}$$

where $P = \mu \restriction \mathcal{B}(\mathcal{T})$ is an ordinary probability measure on the σ-algebra $\mathcal{B}(\mathcal{T})$. Each probability measure on $\mathcal{T}$ is a convex combination of a support measure of the form (2) and a measure of the form (4).

As proved in [22], strict t-norms which are not of the form (3) admit only support measures and no σ-order continuous probability measures.

6 Conditional probability

Conditional probability can be interpreted so that we update our probability model according to partial results. We know that some event occurred and thus we assign a unit probability to it. An even better interpretation may be that the negation of this event *did not occur* and thus it is assigned a zero probability. The new probabilistic model respects this new information and keeps the proportions of probabilities of possible events as much as possible.

Extending conditional probability to fuzzy events is a problem. The excluded middle law does not hold for some choices of operations. Even if it holds, its meaning is different. The occurrence of an event a does not guarantee that a' cannot be observed. Therefore the new (conditional) probability does not have to vanish at a'. This is one of the reasons why the proper definition of conditional probability was formulated as an open problem in [23] (for the special case of an MV-algebra with product, but other models suffer from similar difficulties).

Recent studies of conditional probability on MV-algebras were published in [11, 14, 15].

7 Conclusions

The overlapping of MV-algebras with product and tribes (w.r.t. strict Frank t-norms) shows that two approaches to probability on fuzzy sets converge to essentially the same notions which, moreover, correspond to the original idea by Zadeh. This coincidence opens a new field for further investigations, because some results can be directly translated from one context to the other. Among others, generalizations of the central limit theorem, laws of large numbers [23], and results about entropy can be applied to product (and some other) tribes, too.

On the other hand, a lot of results were derived for tribes, e.g., decomposition theorems, extensions of Lyapunov theorem [1], and applications to games with fuzzy coalitions [3]. These can be applied to MV-algebras with product (at least in the case when they are Łukasiewicz tribes).

References

1. G. Barbieri and H. Weber, "Measures on clans and on MV-algebras," In E. Pap, editor, Handbook of Measure Theory, Elsevier Science, Amsterdam, Chapter 22, pp. 911–946, 2002.
2. G. Barbieri, M. Navara, and H. Weber, "Characterization of T-measures," Soft Computing, vol. 8, pp. 44-50, 2003.
3. D. Butnariu and E. P. Klement, Triangular Norm-Based Measures and Games with Fuzzy Coalitions, Kluwer Academic Publishers, Dordrecht, 1993.
4. D. Butnariu and E.P. Klement, "Triangular norm-based measures," In E. Pap, editor, Handbook of Measure Theory, Elsevier Science, Amsterdam, Chapter 23, 947–1010, 2002.
5. D. Butnariu, E.P. Klement, R. Mesiar, and M. Navara, "Sufficient triangular norms in many-valued logics with standard negation," Arch. Math. Logic, vol. 44, pp. 829–849, 2005.
6. R. Cignoli, I.M.L. D'Ottaviano, and D. Mundici, "Algebraic Foundations of Many-Valued Reasoning," Kluwer Academic Publishers, Dordrecht, 1999.
7. A. Di Nola and M. Navara, "A characterization of σ-complete MV-algebras with enough states," Colloquium Math., vol. 103, pp. 121–130, 2005.
8. A. Dvurečenskij, "Loomis–Sikorski theorem for σ-complete MV-algebras and ℓ-groups," J. Austral. Math. Soc. Ser. A, vol. 68, pp. 261–277, 2000.
9. M. J. Frank, "On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$," Aequationes Math., vol. 19, pp. 194–226, 1979.
10. U. Höhle, "Maße auf unscharfen Mengen," Z. Wahrsch. Verw. Geb., vol. 36, pp. 179–188, 1976.
11. M. Kalina and O. Nánásiová, "Conditional measures and joint distributions on MV-albebras," Proc. 10th Int. Conf. Information Processing and Management of Uncertainty, Perugia, Italy, 2004.
12. E. P. Klement, R. Mesiar, and E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.
13. E. P. Klement, R. Mesiar, and E. Pap, "Triangular norms. Position paper II: general constructions and parameterized families," Fuzzy Sets Syst, vol. 145, pp. 411–438, 2004.
14. T. Kroupa, "Many-dimensional observables on Łukasiewicz tribe: Constructions, conditioning and conditional independence," Kybernetika, vol. 41, no.4, pp. 451–468, 2005.
15. T. Kroupa, "Conditional probability on MV-algebras," Fuzzy Sets and Systems, vol. 149, pp. 369-381, 2005.
16. C. Manara, Relating the Theory of Non-boolean Partitions to the Design of Interpretable Fuzzy Systems, PhD Thesis, Univ. Milano, Italy, 2003.
17. R. Mesiar, "On the structure of T_s-tribes," Tatra Mt. Math. Publ., vol. 3, pp. 167–172, 1993.
18. R. Mesiar, "Nearly Frank t-norms," Tatra Mt. Math. Publ., vol. 16, pp. 127–134, 1999.
19. R. Mesiar and M. Navara, "T_s-tribes and T_s-measures," J. Math. Anal. Appl., vol. 201, pp. 91–102, 1996.
20. M. Navara, "A characterization of triangular norm based tribes," Tatra Mt. Math. Publ., vol. 3, pp. 161–166, 1993.
21. M. Navara, "Nearly Frank t-norms and the characterization of T-measures," In D. Butnariu and E. P. Klement, editors, Proceedings of the 19th Linz Seminar on Fuzzy Set Theory, 9–16, Linz, 1998.
22. M. Navara, "Triangular norms and measures of fuzzy sets," In E. P. Klement and R. Mesiar (eds.), Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms, Elsevier, 345–390, 2005.
23. B. Riečan and D. Mundici, "Probability on MV-algebras," In E. Pap, editor, Handbook of Measure Theory, Elsevier Science, Amsterdam, Chapter 21, 869–910, 2002.
24. S. Weber and E. P. Klement, "Fundamentals of a generalized measure theory," In U. Höhle et al., editors, Mathematics of Fuzzy Sets, Logic, Topology, and Measure Theory, 633–651, Kluwer Academic Publishers, Dordrecht, 1999.
25. R.R. Yager, "Uncertainty representation using fuzzy measures," IEEE Trans. Syst., Man, and Cybernetics, Part B-Cybernetics, vol. 32, pp. 13–20, February 2002.
26. L.A. Zadeh, "Probability measures of fuzzy sets," J. Math. Anal. Appl., vol. 23, pp. 421–427, 1968.

Symmetric Copulas With Given Sub-Diagonal Sections

José Juan Quesada-Molina[1], Susanne Saminger[2,3] and Carlo Sempi[3]

[1] Departamento de Matemática Aplicada
Universidad de Granada
18071 Granada, Spain
jquesada@ugr.es

[2] Department of Knowledge-Based Mathematical Systems
Johannes Kepler University
A-4040 Linz, Austria
susanne.saminger@jku.at

[3] Dipartimento di Matematica "Ennio De Giorgi"
Università di Lecce
I-73100 Lecce, Italy
{susanne.saminger|carlo.sempi}@unile.it

Copulas are functions that join bivariate distribution functions to their univariate marginal distribution functions (see also [7]). In fact, according to Sklars theorem [10], for each random vector (X,Y) there is a copula $C_{X,Y}$ (uniquely defined, whenever X and Y are continuous) such that the joint distribution function $F_{X,Y}$ of (X,Y) may be represented by

$$F_{X,Y}(x,y) = C_{X,Y}(F_X(x), F_Y(y))$$

for all $x,y \in \overline{\mathbb{R}}$, where F_X and F_Y are the distribution functions of the random variables X and Y, respectively. Applications of this fact are provided, e.g., in finance and insurance (see, e.g., [1,3]). Moreover, copulas have played an important role not only in probability theory and statistics, but also in many other fields requiring the aggregation of incoming data such as multicriteria decision making, probabilistic metric spaces and fuzzy theory (see also [4,8,9]).

Therefore, it has been of interest to find methods for constructing copulas by respecting some information given. Among them are particular methods dealing with the fact that one has at his/her disposal some knowledge about the copula along some linear, i.e., horizontal, vertical, or diagonal section. More precisely: Copulas with given diagonal sections have been largely investigated by Fredriks and Nelsen (see also [5]). Those with a given horizontal (or vertical) section have been studied by Klement, Kolesárová, Mesiar and Sempi (see, e.g., [6]). Recently Durante, Kolesárová, Mesiar and Sempi have studied copulas having assigned horizontal *and* vertical sections (see also [2]). In the presentation we will also consider the set of copulas with a given *sub–diagonal* section, namely those copulas that have the same value on the segment of straight line joining the points $(x_0, 0)$ and $(1, 1-x_0)$ with $x_0 \in [0, 1/2]$.

First, by constructing examples, we show that such a set of copulas is not empty. Then we determine the supremum and the infimum of the set considered and show that they are, in fact, copulas, a non–trivial result in view of the known fact that the supremum and the infimum of a set of copulas are quasi–copulas that, in general, need not be copulas. Furthermore, the maximum and the minimum thus found correspond respectively to analogues of the diagonal copula of Fredricks and Nelsen and of the Bertino copula. These latter copulas are respectively the maximum and minimum in the set of diagonal copulas. Finally we show that when one lets x_0 go to zero, the largest and the smallest sub–diagonal copulas tend respectively to the diagonal and the Bertino copulas.

Acknowledgements

This work has been carried out during a post-doc stay of Susanne Saminger at the Dipartimento di Matematica “Ennio De Giorgi” at the University Lecce. She therefore gratefully acknowledges the support by the Austrian Science Fund (FWF) through the Erwin Schrödinger Fellowship ”Dominance – from probabilistic metric spaces to (quasi-)copulas and applications”

References

1. U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. Wiley, Chichester, 2004.
2. F. Durante, A. Kolesárová, R. Mesiar, and C. Sempi. Copulas with given values on a horizontal and a vertical section. *Kybernetika*. (submitted).
3. P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. In S. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier, Amsterdam, 2003.
4. J. C. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.
5. G. A. Fredricks and R. B. Nelsen. Copulas constructed from diagonal sections. pages 129–136. Kluwer Academic Publishers, Dordrecht, 1997.
6. E. P. Klement, A. Kolesárová, R.Mesiar, and C.Sempi. Copulas constructed from horizontal sections. *Comm. Statist. Theory Methods*. (submitted).
7. R. B. Nelsen. *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer, New York, 1999.
8. H. T. Nguyen and E. Walker. *A First Course in Fuzzy Logic*. Chapman & Hall/CRC, Boca Raton, 3rd edition, 2006.
9. B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.
10. A. Sklar. Fonctions de répartition à *n* dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

R-Probability: Fuzzy Probability Idea in Boolean Frame

Dragan Radojević

Mihajlo Pupin Insitute
Volgina 15, 11000 Belgrade, Serbia
Dragan.Radojevic@automatika.imp.bg.ac.yu

1 Introduction

Classical probability [1] is based on the theory of classical sets and, as a consequence, it is in a Boolean frame, (fig 1.). Fuzzy probability [2] is based on the theory of fuzzy sets [3]. Since no algebra of fuzzy sets is a Boolean algebra [4], fuzzy probability is not in a Boolean frame [5]. *Real probability* (***R***-probability) is introduced in this paper. ***R***-probability is based on *Real sets* (***R***-sets). Valued realization of ***R***-sets is based on Interpolative realization of Boolean algebra (IBA) [6], (fig 2. and fig 3.). IBA is a MV realization of finite Boolean algebra, and as a consequence ***R***-probability is a realization of fuzzy probability idea in a Boolean frame.

2 Real sets (R-sets) and Interpolative Boolean algebra

Any element of ***R***-set has analyzed property with *intensity* and/or *gradation* just as a fuzzy set has. ***R***-sets are generated by analyzed ***proper properties***[1]. ***Algebraic - value indifferent characteristics*** of proper properties are: ***Boolean axioms and theorems***. Set of proper properties of interest is ***Boolean algebra of properties*** $\mathrm{BA}p$.

Set of primary proper properties: $\Omega=\{a_1,...,a_n\}$, generates $\mathrm{BA}p(\boldsymbol{\Omega})$ finite Boolean algebra of properties. The basic characteristic of any primary property $a\in\boldsymbol{\Omega}$ is the fact that it can't be represented by the Boolean function of the remaining primary properties. The Boolean algebraic structure of analyzed proper properties is:

$$\langle \mathrm{BA}p(\Omega),\ \cap,\ \cup,\ C\rangle .$$

Any element $\varphi\in\mathbf{BA}p(\boldsymbol{\Omega})$ of finite Boolean algebra of proper properties can be uniquely represented by the following disjunctive canonical form:

$$\varphi=\bigcup_{S\in \mathrm{P}(\Omega)}\sigma_\varphi(S)\alpha(S).$$

Where:

$\alpha(S)=\bigcap_{a_i\in S}a_i\bigcap_{a_j\in\Omega\setminus S}Ca_j$, $(S\in\mathrm{P}(\Omega))$; is ***atomic*** property, atomic element of $\mathrm{BA}p(\Omega)$;

$\sigma_\varphi(S)$, $(S\in P(\Omega))$; is ***structure*** of analyzed property $\varphi\in\mathbf{BA}p(\boldsymbol{\Omega})$.

Structure of any property $\varphi\in\mathrm{BA}p(\Omega)$ is given by the following set function:

$$\sigma_\varphi(S)=\begin{cases}1, & \alpha(S)\subset\varphi\\ 0, & \alpha(S)\not\subset\varphi\end{cases};\ (S\in\mathrm{P}(\Omega)).$$

Fundamental structure's characteristic is the ***principle of structural functionality***[2]: *Structure of any combined property (element of Boolean algebra of analyzed properties) can be directly calculated on the basis of structures of its components by the following rules:*

$$\sigma_{\varphi\cup\psi}(S)=\sigma_\varphi(S)\vee\sigma_\psi(S),$$
$$\sigma_{\varphi\cap\psi}(S)=\sigma_\varphi(S)\wedge\sigma_\psi(S),$$
$$\sigma_{C\varphi}(S)=\neg\sigma_\varphi(S),\quad (S\in\mathrm{P}(S)).$$

[1] or Boolean properties – unary relations

[2] Famous *principle of truth functionality* on value level is only the figure of value irrelevant principle of structural functionality and it is valid only for two-valued case.

Where: $\neg$ unary and $\vee$, $\wedge$ binary classical two-valued Boolean operators.

Structures of primary properties are given by the following set functions:

$$\sigma_{a_i}(S)=\begin{cases}1, & a_i \in S\\ 0, & a_i \notin S\end{cases};\ \left(S\in \mathrm{P}(\Omega),\ a_i\in\Omega\right).$$

Generalized Boolean polynomials uniquely correspond to elements of Boolean algebra of analyzed properties – generators of ***R***-sets. ***R-characteristic function*** of any ***R***-set is its generalized Boolean polynomial.

Atomic R-set $\alpha^{\otimes}(S)$ is the value realization of corresponding atomic property $\alpha(S)$, $(S\in \mathrm{P}(\Omega))$ on analyzed universe of discourses X. ***R***-characteristic function $\alpha^{\otimes}(S):\mathrm{X}\to[0,\ 1]$ of any atomic ***R***-set $\alpha^{\otimes}(S)$, $(S\in \mathbf{P}(\boldsymbol{\Omega}))$ is defined by corresponding ***atomic generalized Boolean polynomial***, [6]:

$$\alpha^{\otimes}(S)(x)=\sum_{K\in \mathrm{P}(\Omega\setminus S)}(-1)^{|K|}\bigotimes_{A_i\in K\cup S}A_i(x),\quad \left(S\in \mathrm{P}(\Omega),\ x\in \mathrm{X}\right).$$

Example: *Atomic Boolean polynomials – atomic **R**-characteristic function for the case when the set of primary properties is* $\Omega=\{a,b\}$ *, are given in the following table:*

S	$\alpha(S)$	$\alpha^{\otimes}(S)(x)$
$\varnothing$	$Ca\cap Cb$	$1-A(x)-B(x)+A(x)\otimes B(x)$
$\{a\}$	$a\cap Cb$	$A(x)-A(x)\otimes B(x)$
$\{b\}$	$Ca\cap b$	$B(x)-A(x)\otimes B(x)$
$\{a,b\}$	$a\cap b$	$A(x)\otimes B(x)$

The intensity of analyzed property $\varphi\in \mathrm{BA}p(\Omega)$ for any element of the universe of discourses $x\in\mathrm{X}$ is given by corresponding ***generalized Boolean polynomial,*** (fig 2.):

$$\varphi^{\otimes}(x)=\sum_{S\in \mathrm{P}(\Omega)}\sigma_{\varphi}(S)\ \alpha^{\otimes}(S)(x),\quad \left(S\in \mathrm{P}(\Omega),\ x\in \mathrm{X}\right).$$

R-characteristic function of $\varphi\in \mathrm{BA}p(\Omega)$ (generalized Boolean polynomial) can be represented as a scalar product of two vectors:

$$\varphi^{\otimes}(x)=\vec{\sigma}_{\varphi}(S)\ \vec{\alpha}^{\otimes}(x).$$

$\vec{\sigma}_{\varphi}=[\sigma_{\varphi}(S)|\ S\in \mathrm{P}(\Omega)]$ and $\vec{\alpha}^{\otimes}(x)=\left[\vec{\alpha}^{\otimes}(S)(x)\ \middle|\ S\in \mathrm{P}(\Omega)\right]$ are *structure vector of* φ and *vector of atomic **R**-characteristic functions*, respectively.

R-set, generated by any property $\varphi\in \mathrm{BA}p(\Omega)$, can be represented as the union of relevant atomic sets:

$$\varphi^{\otimes}=\bigcup_{S\in \mathrm{P}(\Omega)}\sigma_{\varphi}(S)\alpha^{\otimes}(S)\quad .$$

Structures of proper properties preserve the value irrelevant characteristics of ***R***-sets, actually their Boolean nature (Boolean axioms and theorems).

In generalized Boolean polynomials there figure two standard arithmetic operators + and –; and, as a third, a ***generalized product*** $\otimes$ [6]. *Generalized product* is any function $\otimes:[0,\ 1]\times[0,\ 1]\to[0,\ 1]$ that satisfies all four conditions of ***T-norms*** [7]**:** *Commutativity, Associativity, Monotonicity and Boundary conditions* plus one additional: ***Non-negativity condition****:*

$$\sum_{K\in \mathrm{P}(\Omega\setminus S)}(-1)^{|K|}\bigotimes_{a_i\in K\cup S}a_i^{v}\geq 0,\qquad \left(\Omega=\{a_1,..,a_n\},\ S\in \mathrm{P}(\Omega),\ a_i^{v}\in[0,1]\right)$$

The additional axiom “non-negativity” ensures that the values of atomic Boolean polynomials are non-negative: $\alpha^{\otimes}(S)(x)\geq 0,\ \left(S\in \mathbf{P}(\boldsymbol{\Omega}),\ x\in \mathbf{X}\right)$.

***Comment*:** *A generalized product for **R**-sets is just an arithmetic operator; **T**-norm in fuzzy sets has the role of set algebraic operator*.

Example*: *In the case $\Omega=\{a,b\}$ the generalized product, according to the axioms of non-negativity can be in the following interval*[3]*:

$$\max(a+b-1,0)\le a\otimes b\le \min(a,b).$$

R-partition [6] is a consistent generalization of classical sets partition (fig 3.). The collection of atomic ***R***-sets $\left\{ \alpha^{\otimes}(S) \middle| S\in \mathrm{P}(\Omega)\right\}$ is ***R***-partition of analyzed universe of discourses $\mathbf{X}$, since:

(a) atomic sets are *pairwise mutually exclusive***:**

$$\alpha^{\otimes}(S_i)\cap\alpha^{\otimes}(S_j)=\begin{cases}\alpha^{\otimes}(S_i), & i=j\\ \varnothing, & i\neq j\end{cases};\quad \left(\left(\alpha^{\otimes}(S_i)\cap\alpha^{\otimes}(S_j)\right)(x)=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}\right),$$

and (b) they *cover* the universe $\mathbf{X}$:

$$\bigcup_{S\in\mathbf{P}(\Omega)}\alpha^{\otimes}(S)=\mathbf{X};\quad \left(\sum_{S\in\mathbf{P}(\Omega)}\alpha^{\otimes}(S)(x)=1;\quad (x\in\mathbf{X})\right).^{\bullet}$$

In the case of classical sets there is one additional constraint: any element of universe of discourses $x\in\mathbf{X}$ belongs to only one classical atomic set $\alpha(S)$, $(S\in \mathrm{P}(\Omega))$:

$$\alpha(S_i)(x)=1\Rightarrow \alpha(S_j)(x)=0,\ S_j\neq S_i,\ (S_i,S_j\in \mathrm{P}(\Omega))..$$

3. Real probability (*R*-probability)[4]

If a certain event is the universe of discourses X , then any ***R***-set as a subset of the universe of discourses defines a ***random R-event***.

The atomic property $\alpha(S)$ generates on the universe of discourses an atomic ***R***-set $\alpha^{\otimes}(S)$, $(S\in \mathrm{P}(\Omega))$ and/or a random ***R***-event. ***R-probability*** of atomic random ***R***-event $P(\alpha^{\otimes}(S))$ is a mathematical expectation of the ***R***-characteristic function value of a corresponding atomic ***R***-set and/or of the value of a corresponding generalized atomic Boolean polynomial. ***R***-atomic probability in a continual case is given by:

$$P(\alpha^{\otimes}(S))=\int_{x\in X}\alpha^{\otimes}(S)(x)p(x)dx,\quad (S\in \mathrm{P}(\Omega)),$$

and in a discrete case by:

$$P(\alpha^{\otimes}(S))=\sum_{x_i\in X}\alpha^{\otimes}(S)(x_i)\,p(x_i),\quad (S\in \mathrm{P}(\Omega)).$$

$p(x)$ is classic probability distribution function ($\int_{x\in\mathbf{X}}p(x)dx=1$ or $\sum_{x_i\in\mathbf{X}}p(x_i)=1.$).

The sum of ***R***-probabilities of atomic random ***R***-events is identical to 1:

$$\sum_{S\in \mathrm{P}(\Omega)}P(\alpha^{\otimes}(S))=1.$$

since collection of atomic ***R***-events $\left\{\alpha^{\otimes}(S)\ \middle|\ S\in \mathrm{P}(\Omega)\right\}$ is ***R***-partition of universe of discourses (certain event) $\mathbf{X}$.

R-probability of any random ***R***-event is equal to the sum of ***R***-probabilities of atomic random ***R***-events, which are included in it:

$$P(\varphi^{\otimes})=\sum_{S\in\mathbf{P}(\Omega)}\sigma_{\varphi}(S)P(\alpha^{\otimes}(S))$$

Additivity is the principle of ***R***-probability[5] as in the case of classical probability:

---

[3] For $|\Omega|\ge 3$ $\max(a+b-1,0)$ is not low bound of feasible interval for generalized product.

• Element of universe of discourses, in general case can be in more then one atomic ***R***-sets, so that the sum of values of corresponding ***R***-characteristic functions is identically equal to 1.

[4] ***R***-probability is realization of fuzzy probability idea [2] in Boolean frame.

[5] This is not valid in fuzzy probability case.

$$P\left(\bigcup_{S\in \mathbf{P}(\Omega)} \sigma_\varphi(S)\alpha^\otimes(S)\right) = \sum_{S\in \mathbf{P}(\Omega)} \sigma_\varphi(S) P\left(\alpha^\otimes(S)\right).$$

R-probability of any random ***R***-event can be represented as the scalar product of two vectors:

$$P^\otimes\left(\varphi^\otimes\right) = \vec{\sigma}_\varphi \vec{P}\left(\alpha^\otimes\right), \quad \left(\varphi \in BA(\Omega)\right).$$

where: $\vec{P}\left(\alpha^\otimes\right) = \left[P\left(\alpha^\otimes(S)\right) \middle| S \in \mathrm{P}(\Omega)\right]^T$ is *vector of atomic random **R**-event **R**-probabilities* and $\vec{\sigma}_\varphi$ is structural vector of $\varphi \in \mathrm{BA}p(\Omega)$.

The following identities of ***R***-probability are valid for any two $\varphi,\ \psi \in \mathrm{BA}p(\Omega)$ proper properties independently of a chosen generalized product $\otimes$:

$$P\left((\varphi \cup \psi)^\otimes\right) = \left(\vec{\sigma}_\varphi \vee \vec{\sigma}_\psi\right)\vec{P}\left(\alpha^\otimes\right)$$

$$P\left((\varphi \cap \psi)^\otimes\right) = \left(\vec{\sigma}_\varphi \wedge \vec{\sigma}_\psi\right)\vec{P}\left(\alpha^\otimes\right)$$

$$P\left(\left(\varphi^c\right)^\otimes\right) = 1 - P\left(\varphi^\otimes\right)$$

As a consequence, independently of a chosen generalized product and for an arbitrary universe of discourses, all Boolean laws are valid, for example for any $\varphi \in \mathrm{BA}p(\Omega)$:

$$P\left((\varphi \cap C\varphi)^\otimes\right) = 0, \qquad P\left((\varphi \cup C\varphi)^\otimes\right) = 1;$$

$$P\left((\varphi \cap \varphi)^\otimes\right) = P\left(\varphi^\otimes\right), \qquad P\left((\varphi \cup \varphi)^\otimes\right) = P\left(\varphi^\otimes\right).$$

4 Conclusions

In this paper we introduce ***Real probability*** (***R***-probability). ***R***-probability is based on ***Real sets*** (***R***-sets) as classical probability is based on classical sets, (fig 1.). Any element of ***R***-set has generic property by intensity or gradation just as a fuzzy set has. A proper property generates an ***R***-set on the universe of discourses. Value irrelevant characteristics of proper properties are: Boolean axioms and laws. A set of analyzed proper properties is ***Boolean algebra of properties***. Every element of Boolean algebra of properties can be uniquely represented by a ***generalized Boolean polynomial*** (GBP). GBP is an ***R-characteristic function*** of a corresponding ***R***-set, (fig 2). As a consequence all laws of classical set theory are preserved in ***R***-sets and/or by ***R***-sets the idea of fuzzy sets is realized in a Boolean frame, (fig 3). In ***R***-probability a ***random R-event*** is defined by an ***R***-set. The probability of random ***R***-event is the mathematical expectation of its ***R***-characteristic function value. Since both ***R***-sets and classical sets are in the Boolean frame, it follows that ***R***-probability is in the Boolean frame just as classical probability also is.

References

1. A.N. Kolmogorov, Foundations of the theory of probability, Chelsea, reprint, 1950 (Translated from the Russian)
2. L. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl., (23): 421-427, 1968.
3. L. Zadeh, Fuzzy Sets, Information and Control, no. 8, pages 338-353. 1965.
4. P. Hajek, Methamatematics of Fuzzy Logic, Trend in logica – Studia logica library, Kluwer Academic Publishers, Dodrecht/Boston/London, 1988.
5. D.Dubois, H.Prade, Possibility theory, probability and Fuzzy sets: misunderstanding, bridges and gaps, in Fundamentals of Fuzzy sets (Dubois D., Prade H. eds.), The Handbooks of Fuzzy Sets Series, Kluwer, Boston, USA, 2000, 343-438.
6. D.Radojević, Interpolative realization of Boolean algebra as consistent frame for graduation and/or fuzziness, accepted for publication.
7. E.P.Klement, R.Mesiar, E.Pap, Triangular Norms, Kluwer Academic Publishers, Dodrecht, 2000.

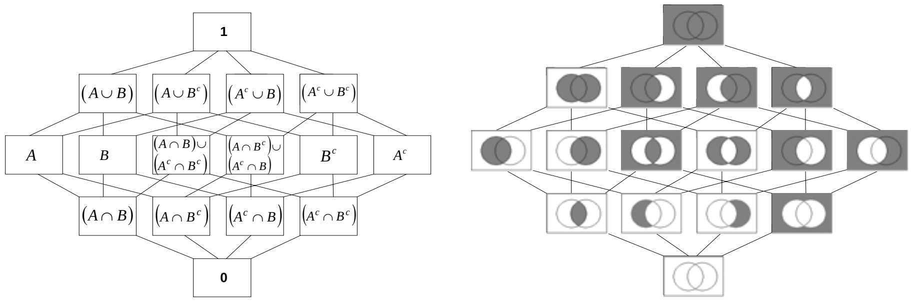

Fig 1: Hasse diagram of Boolean algebra of classical sets generated by two primary sets {*A*, *B*}

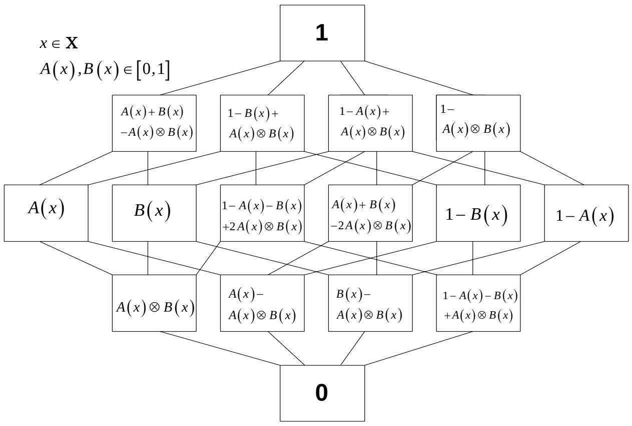

Fig 2: Hasse diagram of ***R***-characteristic functions of Boolean algebra of ***R***-sets generated by two primary ***R***-sets {*A*, *B*}

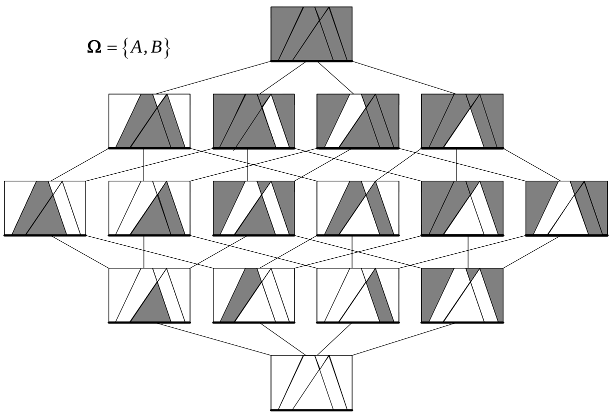

Fig 3: Hasse diagram of Boolean algebra of ***R***-sets generated by two primary ***R***-sets {*A*, *B*} using **min** function as generalized product ⊗

Possibility and Probability on Graph Structures

Arthur Ramer

School of Computer Science and Engineering
University of New South Wales, Sydney 2052, Australia
ramer@cse.unsw.edu.eu

Information theoretic model for possibility distributions is usually based on U-uncertainty. For assignment $\pi = (p_1, \ldots, p_n)$ we first form its descending rearrangement $p_1^{\downarrow} \geq p_2^{\downarrow} \geq \ldots$ and define

$$U(\pi) = \sum (p_i^{\downarrow} - p_{i+1}^{\downarrow}) \log i.$$

This function has been justified, invariably, on the basis of its axiomatic properties. It satisfies all the same 'universal' properties for possibilities as Shannon entropy would for probabilities - additivity, subadditivity and symmetry. Term 'universal' is meant to refer to category theory language where the notions of product, projection and like can be given such a formulation. Related to U-uncertainty is a family of information distances, one of which is a complete metric on the space of all possibility distributions (on a given domain).

Entropy for probabilities has several communication interpretations. In the first part of the paper we introduce a notion of *acceptability* of a choice and propose that U-uncertainty expresses the average complexity of communicating such a choice - it gives the expected length of a minimal necessary message.

In probabilistic modeling, entropy can be extended to recognise the situations when the output symbols are confusable. The most useful is graph entropy proposed by Körner on communication-theoretic grounds. Although the original definition is nonconstructive, it is equivalent to several finitary presentations. It has been used firstly to resolve several questions of zero-error communications. It was later applied to obtain tight lower bounds for sorting of partially ordered sets, both in classical and quantum settings.

The author proposed that it be used to model *imaging* in probability kinematics: it is a generalisation (proposed by several philosophers) to modify conditioning by allowing a nonproportional probability transfer. Given a graph $G = (V, E)$ and probability P on V, $p_i = P(v_i)$, we first look for the set $I = \{I_1, \ldots, I_m\}$ of maximal independent subsets of G. ($I_k \subseteq V$ is independent if it contains no edge from E.) Let Q be a probability distribution on I; we define plausibility of v_i wrt Q

$$\mathrm{Pl}_Q(v_i) = \sum_{I_k \ni v_k} Q(I_k).$$

We now define graph entropy of P wrt G

$$H(P;G) = \min_Q - \sum p_i \log \mathrm{Pl}_Q(v_i).$$

Several instances of belief transfer can be described as minimal change wrt a suitable graph entropies and their associated information divergences. It also serves to resolve successfully the so-called 'Private Benjamin problem', posed by van Fraasen. We discuss these topics in the second part of the paper.

The last part demonstrates how to carry out a similar program for possibility assignments. We extend the notion of U-uncertainty to $U(\pi;G)$, where a graph structure is present on the elements of support of *pi*. Independent sets of G serve to model the (partial) indistinguishability of of the elements, thus of acceptability of choices they represent. We follow by defining graph possibility distance and analyse its properties.

One of the applications is to the question of defining conditional possibility assignment. The author proposed earlier a solution based on the standard U-uncertainty. Other, very attractive methods have been proposed. We show that they can be interpreted as minimal change under a suitable graph possibility.

We close by discussing several open problems and further applications.

Extremes in Nature: An Approach Using Copulas

Gianfautso Salvadori[1] and Carlo De Michele[2]

[1] Dipartimento di Matematica, Università di Lecce
I-73100 Lecce (Italy)
gianfausto.salvadori@unile.it
[2] DIIAR - Sezione CIMI, Politecnico di Milano
I-20133 Milano (Italy)
carlo.demichele@polimi.it

1 Introduction

In many natural problems several are the random variables that play a significant role, and such variates are generally not independent. For instance, the following hydrological examples are paradigmatic of situations that can be found in many geophysical phenomena: different combinations of rainfall intensity and storm duration may generate storms showing quite different characteristics; the river management may strongly depend upon the joint features of flood peak and flood volume; the characterization of droughts requires the joint analysis of duration-magnitude-intensity, and so on.

Therefore, it is often of fundamental importance to be able to link the marginal distributions of different variables in order to obtain a joint law describing the main features of natural events. In general, the development of multivariate probability models (and, in particular, of multivariate Extreme Value distributions) has been limited by mathematical difficulties in generating consistent joint laws with *ad hoc* marginals. Recently, the advent of *Copulas* has solved many of these problems.

In this work we present some recent advances in hydrological modeling which exploit copulas. An application to hydrological data is shown. As a global reference, both on the theoretical and the practical side, the reader is invited to consult [1].

2 The temporal structure of storms

The standard approach to event-based rainfall representation makes a distinction between an "exterior" and an "interior" process. The former one is a coarse representation of rainfall, which characterizes the arrival, duration and average intensity of rainfall events at the synoptic scale. The latter one describes the detailed fluctuations of rainfall intensity at subsynoptic scales. Here we concentrate on the "exterior" process.

The rainfall measurements analysed consist of seven years of hourly rainfall depth measurements collected at the Scoffera station, located in the Bisagno river basin (Thyrrhenian Liguria, Northwestern Italy). We use a non-rainy period lasting (at least) 7 hours to separate two successive storms. In turn, a sequence of 691 storms can be identified and extracted from the data base. For each storm four variables of interest are calculated: (*1*) the storm (average) intensity I (in *mm/h*); (*2*) the storm wet duration W (in *hours*); (*3*) the storm dry duration D (in *hours*) defining the non-rainy period between a storm and the following one; (*4*) the storm volume $V = IW$ (in *mm*). The analysis of the data is carried out studying each of the four seasons separately.

Thus, a storm observation is simply given by the three-components vector (I,W,D). The r.v.'s I,W,D are usually taken as independent, and quite often are given distributions such as Exponential's, Gamma's, or Weibull's: this considerably reduces the mathematical complexity of the model.

However, none of these assumptions are consistent with the data analysed here. In fact, on the one hand a heavy-tailed *Generalized Pareto* (*GP*) law well fits the observed rainfall data I, W, D. On the other hand, a study of the pairwise degree of association between I, W, D makes it evident how these variables are non-independent.

Given the empirical evidence that I, W, D could be non-independent, the problem is how to formalize this fact into a mathematical model. Here we take advantage of the opportunities offered by *Copulas*. Under the (elementary) conditions of Sklar's Theorem, the three bivariate distributions F_{IW}, F_{ID}, and F_{WD} of, respectively, the pairs (I,W), (I,D), and (W,D) can be written in terms of suitable 2-copulas $\mathbf{C}_{IW}$, $\mathbf{C}_{ID}$, and $\mathbf{C}_{WD}$:

$$F_{IW}(i,w) = \mathbf{C}_{IW}\left(F_I(i), F_W(w)\right), \tag{1}$$

$$F_{ID}(i,d) = \mathbf{C}_{ID}\left(F_I(i), F_D(d)\right), \tag{2}$$

$$F_{WD}(w,d) = \mathbf{C}_{WD}\left(F_W(w), F_D(d)\right). \tag{3}$$

Actually, the theory of copulas offers even more possibilities. In fact, the joint law F_{IWD} of the storm vector (I,W,D) can be written as

$$F_{IWD}(i,w,d) = \mathbf{C}_{IWD}\left(F_I(i), F_W(w), F_D(d)\right), \tag{4}$$

where the function $\mathbf{C}_{IWD} : [0,1]^3 \to [0,1]$ represents the 3-copula linking the marginals F_I, F_W, F_D : $\mathbf{R} \to [0,1]$ of the three r.v.'s I, W, D. Evidently, Eq. (4) gives the trivariate distribution function F_{IWD}. In particular, the 3-copula $\mathbf{C}_{IWD}$ could be used to link the three 2-copulas given above:

$$\mathbf{C}_{IWD} \sim \mathbf{C}_{IWD}\left(\mathbf{C}_{IW}, \mathbf{C}_{ID}, \mathbf{C}_{WD}\right). \tag{5}$$

This approach has several advantages. On the one hand, it may be more appropriate to consider a link between bivariate copulas (instead of univariate marginals). On the other hand, the (marginal) stochastic dynamics of the random vectors (I,W) and (W,D) provide the fundamental information for deriving the statistical laws of interest. Note how standard models, involving independent r.v.'s I, W, D, are simple particular cases of the present approach. In fact, consider the 2-copula $\Pi_2(r,s) = rs$ describing pairs of independent r.v.'s. Then, it is enough to replace some of the three 2-copulas in Eq. (5) for Π_2 to obtain the desired model. Furthermore, should I, W, D be fully independent, then the independence 3-copula $\Pi_3(r,s,t) = rst$ could be used directly.

Let us consider the vectors (I,W), (I,D), and (W,D). Several families of 2-copulas were considered to fit the joint distributions F_{IW}, F_{ID}, and F_{WD}. Among many others tested, the 2-copulas belonging to the *Frank*'s family provide a valuable fit to the available data, for all the pairs (I,W), (I,D), (W,D), and the four seasons. Using the values of Kendall's τ estimated on the observed rainfall data, it is easy to write explicitly the distribution functions F_{IW}, F_{ID}, F_{WD} via Eq.s (1)–(3). It is important to point out that such a procedure is distribution-free, since it does not depend upon the knowledge of the marginal laws of I, W, D: therefore, it can be carried out before providing a specific statistical distribution for the variables considered.

A final important point concerns the structure of the 3-copula $\mathbf{C}_{IWD}$. Note that $\mathbf{C}_{IWD}$ represents the mathematical kernel for simulating a sequence of three-components vectors (I,W,D)'s, representing the temporal dynamics of the storms. Among many possible choices, the following function is chosen:

$$\mathbf{C}_{IWD}(r,s,t) = t\,\mathbf{C}_{IW}\left(\frac{\mathbf{C}_{ID}(r,t)}{t}, \frac{\mathbf{C}_{WD}(s,t)}{t}\right), \tag{6}$$

where $r,s,t \in [0,1]$. This 3-copula has a particularly simple and appealing structure, as explained in the following. It is easy to check that $\mathbf{C}_{IWD}$ has three two-dimensional marginals given by the 2-copulas

$\mathbf{C}_{IW}, \mathbf{C}_{ID}, \mathbf{C}_{WD}$. Should any of the three pairs (I,W), (I,D), and (W,D) be formed by independent variables (with 2-copula Π_2), then the expression of $\mathbf{C}_{IWD}$ would further simplify. It is also quite interesting to note how in Eq. (6) the two arguments of $\mathbf{C}_{IW}$ (which rules the joint behavior of (I,W)) are themselves 2-copulas which control, respectively, the pairwise dynamics of (I,D) and (W,D). The parameters of $\mathbf{C}_{IWD}$ are only those of its three marginal 2-copulas $\mathbf{C}_{IW}, \mathbf{C}_{ID}, \mathbf{C}_{WD}$, and hence no further estimations are required to fit $\mathbf{C}_{IWD}$ on the available measurements. In order to check whether this 3-copula is suitable for modeling the available data, a χ^2 test is performed. As a result, an empirically acceptable agreement between the theoretical distribution F_{IWD} given by Eq. (4) and the available observations is found for all seasons.

3 The storm volume

Here we outline how to derive the distribution of the storm volume $V = IW$ by using copulas: a general solution to the problem is given, and we show how to calculate F_V for any suitable 2-copula $\mathbf{C}_{IW}$ and marginals F_I, F_W. Most importantly, these techniques have a broad application in many areas, as illustrated below.

Let us set up the problem in a general framework. Suppose that the distribution function $F_Z(z) = \mathbf{P}\{Z \le z\}$ has to be calculated, where $Z = h(X,Y)$ for some suitable function h. Here the r.v.'s X,Y have continuous and strictly increasing marginals F_X, F_Y and copula $\mathbf{C}_{XY}$. Also, let (R,S) be a random vector, where R,S have *uniform* marginals on $[0,1]$ and joint distribution $F_{RS} \sim \mathbf{C}_{XY}$. Via the *Probability Integral Transform* and the invariance property of copulas, the statistical behavior of (X,Y), and hence of Z, can be modeled as a function of (R,S). In fact:

$$\begin{aligned}\mathbf{P}\{Z \le z\} &= \mathbf{P}\{h(X,Y) \le z\} \\ &= \mathbf{P}\left\{h\left(F_X^{-1}(R), F_Y^{-1}(S)\right) \le z\right\} \\ &= \mathbf{P}\{g(R,S) \le z\},\end{aligned} \tag{7}$$

for some suitable function g. An equivalent formulation can be given in terms of conditional probabilities:

$$\mathbf{P}\{Z \le z\} = \int_0^1 \mathbf{P}\left\{g(R,S) \le z \,\middle|\, S = s\right\} ds, \tag{8}$$

since S is uniform on $[0,1]$. Note that z simply plays the role of a parameter in the integral above. Now, suppose that the inequality $g(R,S) \le z$ can be re-written as $R \le g_z(S)$, for some suitable function g_z (see, e.g., the illustration below). Now, a general result for copulas states that

$$\psi_s(r) = \mathbf{P}\left\{R \le r \,\middle|\, S = s\right\} = \frac{\partial}{\partial s}\mathbf{C}_{XY}(r,s), \tag{9}$$

which exists and is non-decreasing almost everywhere in $[0,1]$. Note that ψ_s is a probability, and therefore it belongs to $[0,1]$. Finally we may write

$$\begin{aligned}F_Z(z) &= \int_0^1 \mathbf{P}\left\{R \le g_z(S) \,\middle|\, S = s\right\} ds \\ &= \int_0^1 \psi_s(g_z(s))\, ds.\end{aligned} \tag{10}$$

As an important result, the calculation of F_Z reduces to a *one-dimensional* integration over $[0,1]$, much easier (theoretically and computationally) than the two-dimensional integration required by standard probabilistic techniques. In some cases the last integral can be solved explicitly, while in general a numerical approach is needed.

As a practical application, let us consider the calculation of the distribution function $F_V(v) = \mathbf{P}\{V \le v\}$ of the storm volume $V = h(I,W) = IW$. For fixed $v > 0$, we may write

$$\begin{aligned}\mathbf{P}\{V \le v\} &= \mathbf{P}\{IW \le v\} \\ &= \mathbf{P}\left\{F_I^{-1}(R)\, F_W^{-1}(S) \le v\right\} \\ &= \mathbf{P}\{R \le g_v(S)\}\,, \end{aligned} \tag{11}$$

where the function

$$g_v(s) = F_I\left(v / F_W^{-1}(s)\right) \tag{12}$$

is monotonous and continuous on the compact subset $[0,1]$, ranges in $[0,1]$, and is bounded and integrable w.r.t. s on the unit interval. Then, F_V can be calculated via Eq. (10), simply by replacing Z for V and $\mathbf{C}_{XY}$ for $\mathbf{C}_{IW}$. Considering the rainfall measurements collected at the Scoffera station, the comparison between the empirical distribution F_V and the corresponding theoretical expression (as calculated above) shows that in all cases the agreement is impressive over all the wide range considered — for the available data V takes on values in the interval 0–500 mm.

References

1. G. Salvadori, C. De Michele, N.T. Kottegoda, and R. Rosso. *Extremes in nature. An approach using copulas*. Springer-Verlag, 2007.

Triangle Functions — a Way to Bridge Gaps?

Susanne Saminger[1,2] and Carlo Sempi[2]

[1] Department of Knowledge-Based Mathematical Systems
Johannes Kepler University
A-4040 Linz, Austria
susanne.saminger@jku.at
[2] Dipartimento di Matematica "Ennio De Giorgi"
Università di Lecce
I-73100 Lecce, Italy
{susanne.saminger|carlo.sempi}@unile.it

This year's Linz Seminar on Fuzzy Set Theory is entitled "Fuzzy Sets, Probabilities, and Statistics — Gaps and Bridges". It is remarkable that since the introduction of statistical, later on, called probabilistic metric spaces, by Menger (see [5], but also [7, 8]), the investigations in these fields yielded particular operations and important results which are related to at least two of the aspects mentioned before — namely fuzzy sets and probabilities. Whereas the first one is inevitably connected to triangular norms, an indispensable tool for modelling their conjunction in this many-valued framework (see also [1–4]), the second field relates to copulas (see also [6]). A class of operations which are essential, due to Sklar's theorem [9], in the fields of joint distribution functions with given marginals and completely bearing the dependence structure of the related underlying margins.

However, probabilistic metric spaces are also inconceivable without another class of operations, namely triangle functions. They are particularly important in the generalization of the triangle inequality of metric spaces to probabilistic metric spaces and are as such an operation on the set of (distance) distribution functions. Moreover, several types of triangle functions relate to triangular norms and copulas and as such build another bridge between those two areas.

A *triangle function* is a binary operation on the set Δ^+ of distributions functions F satisfying the condition $F(0) = 0$, that is commutative, associative, and non-decreasing in each place and has ε_0 as identity. Explicitly a triangle function τ satisfies the following conditions, for all F, G and H in Δ^+:

(TF1) $\tau(\tau(F,G),H) = \tau(F,\tau(G,H))$;
(TF2) $\tau(F,G) = \tau(G,F)$;
(TF3) if $F \leq G$, then both $\tau(F,H) \leq \tau(G,H)$ and $\tau(H,F) \leq \tau(H,G)$;
(TF4) $\tau(\varepsilon_0,F) = \tau(F,\varepsilon_0) = F$.

Particularly, we will focus on the following various families of triangle functions:

- the pointwise induced triangle functions; and we study also the connection with aggregation operators;
- the triangle functions of the type $\tau_{T,L}$:

$$\tau_{T,L}(F,G)(x) = \sup\{T(F(u),G(v)) \mid L(u,v) = x\},$$

  where T is a t-norm and L a binary operation on $\mathbb{R}_+$;
- the triangle functions of the type $\tau_{T^*,L}$:

$$\tau_{T^*,L}(F,G)(x) = \inf\{T^*(F(u),G(v)) \mid L(u,v) = x\},$$

  where T^* is the t-conorm of a t-norm T;

– the triangle functions of the type $\sigma_{C,L}$:

$$\sigma_{C,L}(F,G)(x) := \int_{\{(u,v):L(u,v)<x\}} dC(F(u),G(v));$$

where C is a copula belonging to a subset characterized by Frank. These triangle functions include the classical convolution of distribution functions and have a clear probabilistic meaning.

– the triangle functions of the type $\rho_{Q,L}$:

$$\rho_{Q,L}(F,G)(x) = \inf\{\overline{Q}(F(u),G(v)) \mid L(u,v) = x\}$$

where Q is a quasi-copula and $\overline{Q}$ is defined by $\overline{Q}(x,y) := x+y-Q(x,y)$.

The aim of the presentation is therefore not just to go back to the roots and to recall the very basic concepts but in particular to reconceive them in light of recent results in the fields of triangular norms, (quasi-)copulas as well as aggregation operators and as such provide new perspectives for future investigations.

Acknowledgements

This work has been carried out during a post-doc stay of Susanne Saminger at the Dipartimento di Matematica “Ennio De Giorgi” at the University Lecce in the framework of the Erwin Schrödinger Fellowship ”Dominance – from probabilistic metric spaces to (quasi-)copulas and applications” sponsored by the Austrian Science Fund (FWF). The support is gratefully acknowledged.

References

1. C. Alsina, M. Frank, and B. Schweizer. *Associative Functions: Triangular Norms and Copulas*. World Scientific Publishing Company, Singapore, 2006.
2. P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.
3. E. P. Klement and R. Mesiar, editors. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, Amsterdam, 2005.
4. E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic. Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.
5. K. Menger. Statistical metrics. *Proc. Nat. Acad. Sci. U.S.A.*, 8:535–537, 1942.
6. R. B. Nelsen. *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer, New York, 1999.
7. B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.
8. B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. Dover Publications, Mineola, 2nd edition, 2005.
9. A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

Fuzzy Inclusion and Similarity Through Coherent Conditional Probability

Romano Scozzafava and Barbara Vantaggi

Dept. Metodi e Modelli Matematici
Università "La Sapienza", Rome, Italy
{romscozz|vantaggi}@dmmm.uniroma1.it

1 Introduction

In the framework of the interpretation of fuzzy theory in terms of coherent conditional probability, as given by Coletti & Scozzafava (see, e.g., [3, 4]), we deal with binary fuzzy relations (e.g., inclusion, similarity, etc.). Some preliminary results have been presented in [2].

Consider the following (classic) example: if X is a numerical quantity and φ_X is the property "small", for You the membership function $\mu(x)$ may be put equal to 1 for values x of X less than a given x_1, while it is put equal to 0 for values greater than x_2; then it is taken as decreasing from 1 to 0 in the interval from x_1 to x_2. This choice of the membership function implies that, for You, elements of the range C_X of X less than x_1 have the property φ_X, while those greater than x_2 do not. So the real problem is that You are doubtful (and then uncertain) on having or not the property φ_X those elements of C_X between x_1 and x_2. Then the interest is in fact directed toward *conditional events* such as $E_\varphi|A_x$, where $A_x = \{X = x\}$, and x ranges over the interval from x_1 to x_2, with

$$E_\varphi=\{\text{You claim (that } X \text{ has) the property } \varphi_X\}.$$

In other words, we identify the values of the membership function $\mu(x)$ with suitable conditional probabilities. Notice that this conditional probability $P(E_\varphi|A_x)$ is *directly* introduced as a function on a set of conditional events (and without assuming any given algebraic structure). Is that possible? In the usual (Kolmorogovian) approach to conditional probability the answer is NO, since the introduction of $P(E_\varphi|A_x)$ would require the consideration (and the assessment) of $P(E_\varphi \wedge A_x)$ and $P(A_x)$ (assuming positivity of the latter).

These problems are easily by–passed in our framework based on coherent conditional probability. For the sake of brevity, we just mention that a coherent conditional probability on an arbitrary family $\mathcal{C}$ can be characterized by suitably representing it (in any finite subset $\mathcal{F}$ of $\mathcal{C}$) by means of a *class* $\{P_\alpha\}$ of coherent unconditional probabilities giving rise to the so-called *zero-layers* (indexed by α): we refer to [4] (p.81) for this theory.

For the formal definitions concerning fuzzy sets through coherent conditional probability, see [3]. We are able not only to define fuzzy subsets, but also to introduce in a very natural way the counterparts of the basic continuous T-norms and the corresponding dual T-conorms, bound to the former by *coherence*. In this framework, we will introduce some specific binary fuzzy relations.

2 Fuzzy Inclusion

Let $\mathcal{F}(\mathcal{C}_X)$ be the family of fuzzy subsets, i.e. the set of pairs $\{E_\pi, \mu_{E_\pi}\}$ where μ_{E_π} is defined on $\mathcal{C}_X$ and $\mu_{E_\pi}(\cdot) = P(E_\pi|\cdot)$ is a coherent conditional probability.

Consider now any two fuzzy subsets $E^*_\varphi = (E_\varphi, \mu_{E_\varphi})$, $E^*_\psi = (E_\psi, \mu_{E_\psi})$ of $\mathcal{C}_X$. Thus, since the assessment $P(\cdot|\cdot)$ defined on the following set of conditional events $\mathcal{C} = \{E_\varphi|A_x\,,\, E_\psi|A_x \,:\, A_x \in C_X\}$ is coherent, it can be extended (preserving coherence) to any set $\mathcal{D} \supset \mathcal{C}$.

So we define *degree of fuzzy inclusion* a function

$$I : \mathcal{F}(\mathcal{C}_X) \times \mathcal{F}(\mathcal{C}_X) \to [0,1]$$

with

$$I(E^*_\varphi, E^*_\psi) = P\big(E_\psi | (E_\varphi \vee E_\psi)\big)\,,$$

obtained as any coherent extension of $P(\cdot|\cdot)$ to the conditional event $E_\psi|(E_\varphi \vee E_\psi)$.

The existence of such a function is warranted by a fundamental extension Theorem for coherent conditional probabilities ([4]). The semantic behind this choice is the following: "the more" E^*_φ is included in E^*_ψ, "the more" if You claim *at least one* of the two corresponding properties You are willing to claim the property ψ.

In the case of crisp sets we obtain that fuzzy inclusion holds with degree 1: in fact, if $A \subseteq B$, then $P(B|A \vee B) = P(B|B) = 1$. However, even in the crisp case, the fact that $P(E|E \vee F) = 1$ does not imply that $F \subseteq E$. Notice that in the crisp case we have that inclusion is reflexive, i.e. any set A is such that $A \subseteq A$.

As far as fuzzy inclusion is concerned, we have that any fuzzy set E^*_ψ is included in itself with degree $I(E^*_\psi, E^*_\psi) = 1$, so also fuzzy inclusion is necessarily reflexive.

An interesting property of the degree of fuzzy inclusion of E^*_ψ in E^*_φ (and that of inclusion of E^*_φ in E^*_ψ) is given by the following inequality

$$I(E^*_\psi, E^*_\varphi) \geq 1 - I(E^*_\varphi, E^*_\psi), \tag{1}$$

an easy consequence of an elementary property of conditional probability.

To compute $I(E^*_\psi, E^*_\varphi)$, notice that, given $\mu_\varphi(\cdot) = P(E_\varphi|\cdot)$ and $\mu_\psi(\cdot) = P(E_\psi|\cdot)$ defined on C_X, we can find also the membership function $\mu_{\varphi\cup\psi}(\cdot)$ of $(E^*_\psi \cup E^*_\varphi)$ as coherent extension of the assessment P given on $\{E_\psi|A_x\,,\, E_\varphi|A_x \,:\, A_x \in C_X\}$ (corresponding to a t-conorm: see [3]). Then, given a conditional probability $P(\cdot|\cdot)$ on $\mathcal{A}_X \times \mathcal{A}^o_X$ (with $\mathcal{A}_X$ the algebra generated by the events A_x and $\mathcal{A}^o_X = \mathcal{A}_X \setminus \{\emptyset\}$), it gives rise to a class $\{P_\alpha\}$ of coherent (unconditional) probabilities, and so (for simplicity we refer to a finite $\mathcal{C}_X$)

$$I(E^*_\psi, E^*_\varphi) = \frac{\sum_x \mu_\varphi(x) P_\alpha(x)}{\sum_x \mu_{\varphi\cup\psi}(x) P_\alpha(x)}\,,$$

where α is the zero-layer of the event $E_\psi \vee E_\varphi$ (i.e., it is such that the denominator of the above fraction is strictly greater than zero, see [4]).

A possible requirement for *inclusion* could be some (weak) form of transitivity (many definitions in the relevant literature lack this property). In fact, a strong form of transitivity, based on the minimum t-norm (and called min-transitivity), requires that for any $E^*_\psi, E^*_\varphi, E^*_\nu \in \mathcal{F}(\mathcal{C}_X)$

$$I(E^*_\psi, E^*_\nu) \geq \min\big\{I(E^*_\psi, E^*_\varphi), I(E^*_\varphi, E^*_\nu)\big\}$$

and different authors maintain that it is a too strong requirement: our definition of degree of fuzzy inclusion does not necessarily satisfy min-transitivity, while a weaker form of transitivity, Łukasiewicz-transitivity, holds, in fact for any E^*_φ,E^*_ψ,$E^*_\nu \in \mathcal{L}(C_X)$

$$I(E^*_\varphi, E^*_\psi) \geq \max\{I(E^*_\varphi, E^*_\nu) + I(E^*_\nu, E^*_\psi) - 1, 0\}\,.$$

Then in our setting the degree of fuzzy inclusion can be called *weakly transitive*.

3 Similarity

In [2] similarity has been introduced, on the basis of the interpretation of fuzzy sets within the theory of coherent conditional probabilities, in the following way.

Let $\mathcal{F}(\mathcal{C}_X)$ be the family of fuzzy subsets E^*_π of $\mathcal{C}_X$. A *similarity* S is a mapping

$$S : \mathcal{F}(\mathcal{C}_X) \times \mathcal{F}(\mathcal{C}_X) \longrightarrow [0,1]$$

satisfying

1. (Symmetry) $S(E^*_\varphi, E^*_\psi) = S(E^*_\psi, E^*_\varphi)$;
2. (Reflexivity) $S(E^*_\varphi, E^*_\varphi) = 1$.

Then, given any two fuzzy subsets $E^*_\varphi = (E_\varphi, \mu_{E_\varphi})$, $E^*_\psi = (E_\psi, \mu_{E_\psi})$ of $\mathcal{C}_X$, with $\mu_{E_\varphi}(\cdot) = P(E_\varphi|\cdot)$ (and analogously for ψ), let $P\big((E_\varphi \wedge E_\psi|A_x)\big)$ be a relevant coherent assessment (corresponding to a "T-norm") and the associate coherent assessment $P\big((E_\varphi \vee E_\psi)|A_x\big)$ (corresponding to a dual t-conorm). Any coherent extension of $P(\cdot|\cdot)$ to the conditional event $(E_\varphi \wedge E_\psi)|(E_\varphi \vee E_\psi)$ is a similarity, i.e.

$$S(E^*_\varphi, E^*_\psi) = P\big((E_\varphi \wedge E_\psi)|(E_\varphi \vee E_\psi)\big) .$$

The existence of such a function is warranted by the aforementioned extension Theorem for coherent conditional probabilities.

The semantic behind this choice is the following: "the more" two fuzzy subsets are considered to be similar, "the more" if You claim *at least one* of the two corresponding properties You are willing to claim *both* properties.

Now we show how to compute $S(E^*_\varphi, E^*_\psi)$: given $\mu_\varphi(\cdot) = P(E_\varphi|\cdot)$ and $\mu_\psi(\cdot) = P(E_\psi|\cdot)$, the membership functions $\mu_{\varphi\cup\psi}(\cdot)$ and $\mu_{\varphi\cap\psi}(\cdot)$ of the fuzzy sets $(E^*_\psi \cup E^*_\varphi)$ and $(E^*_\psi \cap E^*_\varphi)$ (corresponding to a t-conorm and a dual t-norm, see [3]) arise as coherent extensions of the assessment P given on $\{E_\psi|A_x\,,\, E_\varphi|A_x \,:\, A_x \in C_X\}$.
Given a conditional probability $P(\cdot|\cdot)$ on $\mathcal{A}_X \times \mathcal{A}^o_X$ (which gives rise to a class $\{P_\alpha\}$ of coherent probabilities), we have (for simplicity we refer to a finite $\mathcal{C}_X$)

$$S(E^*_\varphi, E^*_\varphi) = \frac{\sum_x \mu_{\varphi\cap\psi}(x)\lambda_\alpha(x)}{\sum_x \mu_{\varphi\cup\psi}(x)\lambda_\alpha(x)}$$

where $\lambda_\alpha(x) = P_\alpha(A_x)$, with α the zero-layer of the event $E_\psi \vee E_\varphi$.

Notice that (contrary to what happens in the classic fuzzy framework) this approach to similarity is able to take into account – through the probability values $\lambda_\alpha(x)$ – possible different "weights" of the values x.

We show now how some classic similarity functions (the most used in applications and proposed in the relevant literature, see, e.g. [1, 5]) are related to the above formula involving conditional probability.

By choosing as t-norm the minimum T_M, we get

$$S(E^*_\varphi, E^*_\psi) = \frac{\sum_x \min\big\{P(E_\varphi|A_x), P(E_\psi|A_x)\big\}\lambda_\alpha(x)}{\sum_x \max\big\{P(E_\varphi|A_x), P(E_\psi|A_x)\big\}\lambda_\alpha(x)} .$$

For suitable choices of the probabilities $\lambda_\alpha(x)$, the classic similarity functions are obtained. For example, taking a constant probability $\lambda_\alpha(x)$, we obtain

$$S(E^*_\varphi, E^*_\psi) = \frac{\sum_x \min\big\{P(E_\varphi|A_x), P(E_\psi|A_x)\big\}}{\sum_x \max\big\{P(E_\varphi|A_x), P(E_\psi|A_x)\big\}} = \frac{\sum_x \min\big\{\mu_\varphi(x), \mu_\psi(x)\big\}}{\sum_x \max\big\{\mu_\varphi(x), \mu_\psi(x)\big\}} ,$$

which is a well–known similarity function.

The similarity S does not generally satisfy the T-transitivity property for some suitable t-norm (as $T = T_M$), however, the similarity S satisfies the T-transitivity property for T=T_L (Łukasiewicz T-norm), i.e.

$$S(E^*_\varphi, E^*_\psi) \geq \max\{0\,,\; S(E^*_\varphi, E^*_\nu) + S(E^*_\nu, E^*_\psi - 1\}$$

for any $E^*_\varphi, E^*_\psi, E^*_\nu \in \mathcal{L}(C_X)$.

The relationship between the degree of fuzzy inclusion and similarity is given through the following equality $S(E^*_\varphi, E^*_\psi) = I(E^*_\varphi, E^*_\psi) + I(E^*_\psi, E^*_\varphi) - 1$, which implies that the degree of similarity between two fuzzy sets E^*_φ and E^*_ψ depends on the degrees of the fuzzy inclusion of E^*_φ in E^*_ψ and that of E^*_ψ in E^*_φ. Thus, if the degree γ of inclusion of E^*_φ in E^*_ψ and that of E^*_ψ in E^*_φ are equal (in this case eq. (1) implies that $\gamma \geq \frac{1}{2}$), it follows that "the more" E^*_φ is included in E^*_ψ and E^*_ψ is included in E^*_φ, "the more" the two fuzzy sets E_φ and E_ψ are similar.

References

1. B. Bouchon-Menieur, M. Rifqi, S. Bothorel (1996) Towards general measures of comparison of objects. *Fuzzy Sets and Systems* 84, 143–153.
2. R. Scozzafava, B. Vantaggi (2006) Fuzzy relations in a coherent conditional probability setting. *7th International Conference on Information and Management Sciences* (IMS), Chengdu (China), pp. 496-500.
3. G. Coletti, R. Scozzafava (2004) Conditional Probability, Fuzzy Sets, and Possibility: a Unifying View. *Fuzzy Sets and Systems* 144, 227–249.
4. G. Coletti, R. Scozzafava (2002) *Probabilistic logic in a coherent setting*. (Trends in logic n.15) Kluwer, Dordrecht.
5. D. Dubois, H. Prade (1980) *Fuzzy sets and systems, theory and applications*. Academic Press, New-York.

From Observable-Based to Probabilistic Peasoning

Claudio Sossai

ISIB-CNR
Padova, Italy
claudio.sossai@isib.cnr.it

In this work, using categorical techniques, I will give a mathematical definition of law of chance. I will also show that every proof in the multiplicative fragment of linear logic can be interpreted in a law of chance (validity). Laws of chance are defined as time and uncertainty invariants. I believe that they can give an interesting contribution to answer the following question: why is mathematics reliable? It is a common opinion that even a partial answer to this question could give some insight to the problem of the foundations of mathematics. There are many examples of the reliability of the mathematical method in different theories and fields: for instance the existence of the planet Pluto has been foreseen only on the basis of mathematical computations. Using the validity of the proof system the reliability of reasoning (and I believe also of computing due to the Curry-Howard isomorphism) is a consequence of the fact that these methods are based on the laws of chance. Such laws are satisfied by many possible outcomes that have not yet been observed. In fact proofs, in this semantics, define infinite sets of possible observables, while the available information is only finite. My claim is therefore that mathematics is reliable because it is able to grasp some of these invariants that remain stable also in the presence of the high variability of outcomes due to randomness. This aspect gives us the possibility of defining non local rules (the ones of logic and computations) used to give meaning to local observations (the ones available to us), i.e. rules that allow us to forecast what we have not yet observed, like in the example of the discovery of Pluto.

A filtration is a family $\{\mathcal{F}_n\}_{n\in N}$ of subalgebras of $\mathcal{B}$ (Boolean algebra) s.t. $\mathcal{F}_n \subseteq \mathcal{F}_m$ if $n \leq m$. $\{\mathcal{F}_n\}_{n\in N}$ is adapted for a stochastic process $\{X_n\}_{n\in N}$ iff X_n is $\mathcal{F}_n$-measurable for every n. A trajectory, or observable, of $\{X_n\}_{n\in N}$ is a finite set of outcomes: $x_n(\xi) = \{X_i(\xi) : i \leq n\}$. It is a well known fact that all information contained in a stochastic process is described by its filtration. For this reason I will assume that two stochastic processes that generate the same filtration are equivalent.

Definition 1. *An experiment is defined as a stochastic process $X = \{X_n\}_{n\in N}$ with a method, σ^X, that associates to every trajectory $x_n(\xi)$ a measure $\sigma^X_{x_n(\xi)}$ defined over $\mathcal{F}_0$.*

I will indicate with $|X|$ the set of all observables of X. For example in the case of independent processes, the Glivenko-Catelli theorem gives a sound method to define experiments.

Experiments X and Y can be combined giving a compound experiment, here called product experiment $X \times Y$. The idea is that the filtration associated to $X \times Y$ is the intersection of the two filtrations, i.e. $\mathcal{F}^{X\times Y} = \{\mathcal{F}^X_n \cap \mathcal{F}^Y_n\}_{n\in N}$ and the method $\sigma^{X\times Y}$ is defined using the product measure of σ^X and σ^Y. Therefore X and Y are assumed to be independent experiments. An observable of $X \times Y$ can be described as a pair xy where x,y are observables of X and Y respectively. For the technical details see [7]. If $x \in |X|$ then n is length of x indicated with $l(x)$. Note that if $xy \in |X \times Y|$ then $l(x) = l(y)$. The logical language (multiplicative fragment with exponentials of Linear Logic [1]) is: $L := P|\perp|1|\phi^\perp|\phi\otimes\psi|\phi\parr\psi|?\phi|!\phi$. To give a stochastic semantics to logic the following is assumed: to every atomic proposition P it is associated an experiment (indicated with P); the experiment associated to $\phi^\perp$ is the same as the one associated to ϕ; the experiment $\phi\times\psi$ is associated to $\phi\otimes\psi$ and $\phi\parr\psi$. The connectives are defined using the associated experiment and a relation of coherence (see below) between the observables of the experiment. In statistics, a test is a method that is used to exclude a measure given a set of data. An example is the use of observations to test the efficiency of a treatment in medicine. To this aim, the observations are used to exclude the measure that describes the fact that the difference between the two expected values of the populations has mean 0 (usually called hypothesis H_0). Given two observations x and x' of the experiment P, there are only two possibilities: **1)** x' can be used to confirm the empirical distribution of x, i.e. $\sigma_x^P = \sigma_{x'}^P$. In this case I call x' a statistic for x (shortly $x'\ STAT_P\ x$), **2)** x' can be used as a test against the empirical distribution of x, i.e. $\sigma_x^P \neq \sigma_{x'}^P$. In this case x' is a test against x (shortly $x'\ TEST_P\ x$). Note that $xSTAT_Px'$ iff **not** $xTEST_Px'$. Using these ideas negation can be defined as: $xSTAT_{P^\perp}x'$ iff $\neg xSTAT_Px' \vee x = x'$ where $x = x'$ is required to preserve the fact that $STAT$ is a symmetric relation. To see how the duality between$TEST$ and $STAT$, in compound experiments, generates the connectives, let me recall two properties of product measures. 1) If $\mu = \lambda \times \nu$ and $\mu' = \lambda' \times \nu'$ then $\mu = \mu'$ iff $\lambda = \lambda'$ **and** $\nu = \nu'$, therefore it also holds that $\mu \neq \mu'$ iff $\lambda \neq \lambda'$ **or** $\nu \neq \nu'$. 2) A product measure $\mu \times \nu$ defined over $\mathcal{B}_1 \times \mathcal{B}_2$ can induce a measure $(\mu \times \nu)_1$ over $\mathcal{B}_1$ or a measure $(\mu \times \nu)_2$ over $\mathcal{B}_2$. Using these properties, we can see that in a compound experiment the $STAT$ relation generates the (positive) connective $\otimes$ (that behaves like an **and**) while the $TEST$ relation generates a (negative) connective $\parr$ (that behaves like an **or**). In fact, using the above properties, we have that: $xySTAT_{P\otimes Q}x'y'$ iff $\sigma_{xy}^{P\times Q} = \sigma_{x'y'}^{P\times Q}$ iff $(\sigma_{xy}^{P\times Q})_1 = (\sigma_{x'y'}^{P\times Q})_1$ **and** $(\sigma_{xy}^{P\times Q})_2 = (\sigma_{x'y'}^{P\times Q})_2$ iff $xSTAT_Px'$ **and** $ySTAT_Qy'$. Summing up from the above equations we obtain the definition of the $\otimes$ connective: $xySTAT_{P\otimes Q}x'y'$ iff $xSTAT_Px'$**and** $ySTAT_Qy'$ If we define, as usual, $P\parr Q = (P^\perp \otimes Q^\perp)^\perp$ then from the duality $STAT/TEST$ we obtain the second connective (or): $xySTAT_{(P\parr Q)}x'y' \wedge xy \neq x'y'$ iff $(xSTAT_Px' \wedge x \neq x')$ **or** $(ySTAT_Qy' \wedge y \neq y')$. In [7] it is proved that the $STAT$ relation satisfies the properties that define the coherent relations in the denotational semantics of linear logic. Therefore it is possible to prove the validity theorem i.e. every proof π of a formula ϕ can be interpreted in a set π^* of coherent observables (called clique), i.e. for every pairs $x, x' \in \pi^*$ it holds that $xSTAT_\phi x'$.

What is the meaning of this result?

The claim is that proofs, in this semantics, are chance and time invariants. To prove this claim we must first give a mathematical definition of law of chance (that will be defined as a true formula in a suitable topos of presheaves) and then show that every proof can be interpreted in a law of chance. Let $X_1, ...X_n$ be the set of available experiments and $X = X_1 \times ... \times X_n$ with $x \in |X|$.

Definition 2. *The category $\mathcal{T}$ of time is defined as follows:*

- *the objects of $\mathcal{T}$ are the sets $A_{x_i} = \{\sigma_x^X : \sigma_x^X(i_i(B_i)) = \sigma_{x_i}^{X_i} \wedge l(x) \geq l(x_i)\}$, where $x_i \in |X_i|$*
- *there is an arrow $f : A_{y_i} \to A_{x_i}$ iff $x_i, y_i \in |X_i|$, $l(y_i) \geq l(x_i)$ and $\sigma_{x_i}^{X_i} = \sigma_{y_i}^{X_i}$.*

where $i_i : \mathcal{B}_i \to \mathcal{B}$ is the immersion of B_i in $\mathcal{B} = \Pi_{i \leq n}\mathcal{B}_i$, i.e. what in measure theory is called product algebra (see [2]), indeed a co-product. Note the following simple fact: if $f : B \to A$ is an arrow of $\mathcal{T}$ then $B \subseteq A$. In this mathematical framework every observable x_i can be seen as a time point, where two observables x_i, y_i characterize the same time point iff $l(x_i) = l(y_i)$ and $\sigma_{x_i}^{X_i} = \sigma_{y_i}^{X_i}$. An arrow of the category $\mathcal{T}$ links two observables x_i and y_i where y_i comes after x_i and the associated statistical measures are equal. Here time has a more complicated structure w.r.t. the usual linear order of events. This is due to the fact that this idea of time contains that of uncertainty. In fact the description of uncertainty will stem as a natural consequence from a mathematical construction built on the category of time, i.e. the topos of presheaves over $\mathcal{T}$, where, as we will see, uncertainty will be described in a natural way, without any ad hoc hypothesis. Nonetheless $\mathcal{T}$ contains, as a special case, the usual definition of time. To see this, let $1 = \{1_n\}_{n\in N}$ with $1_n(\xi) = 1$ for all n and $\xi \in \Xi$. Then the filtration associated has all elements equal to $\{\Xi, \emptyset\}$, the elements $n \in |1|$ have the form $n = \langle 1, 1, ...1\rangle$ n-times and the statistical measure is $\sigma_n^1(\Xi) = 1$. It is easy to see that $1 \times X = X$ for every experiment and if $|1|$ is the coherent space associated to 1 then $|1| \otimes |X| \simeq |X|$. Note that 1 is what we usually call time or better linear time, measured by clocks. In fact a clock is nothing but an experiment that has only one possible observable (i.e. time passing) that surely happens (till uncertainty comes into play i.e. the clock, as a machine, breaks). Note that linear time is the witness of provability, in fact completeness of linear logic w.r.t. phase semantics reads: $\vdash \phi$ iff $1 \in ||\phi||$.

Definition 3. *$\mathcal{T}$–Sets is the topos of presheaves over $\mathcal{T}$*

It is interesting to note that $\mathcal{T}$–*Sets* contains a categorical definition of measure. To see this, note that $\mathcal{T}$–*Sets* has the following object, i.e. the contravariant functor $\mathcal{R} : \mathcal{T}^{op} \to Sets$, defined as:

1. $\mathcal{R}(A) = \{k : A \xrightarrow{k} R\}$, i.e., the set of all functions k from A to the set R of the usual real numbers,
2. for $B \xrightarrow{f} A$, $\mathcal{R}(A) \xrightarrow{\mathcal{R}(f)} \mathcal{R}(B)$ is the function that sends every $g \in \mathcal{R}(A)$ to g restricted to B.

Useful objects of $\mathcal{T}$–*Sets* are the *constant functors* Δ^S (where S is a set) defined as $\Delta^S(A) = S$ for every object A and $\Delta^S(f) = 1_S$ (identity map) for every arrow f. In particular, we will consider $\Delta^{\mathcal{B}}$, where $\mathcal{B}$ is the Boolean algebra, and Δ^R, where R is the set of real numbers. Using the internal language of $\mathcal{T}$–*Sets*, it is possible to prove that the elements of the constant presheaf Δ^R satisfy the properties of Dedekind cuts [3], hence we will take Δ^R as the set of real numbers in $\mathcal{T}$–*Sets*. Note that Δ^R can be embedded in $\mathcal{R}$. In fact, for every usual real number $r \in R$, let $\ulcorner r \urcorner \in \mathcal{R}(A)$ be the function that sends A to r, i.e., for every $\sigma^X_x \in A$, $\ulcorner r \urcorner : \sigma^X_x \longmapsto r$, then $\iota^{\Delta^R} : \Delta^R \to \mathcal{R}$ is defined, for all A, $r \in \Delta^R(A)$, as $: \iota^{\Delta^R}_A(r) = \ulcorner r \urcorner$. When the context is clear, we will write r for the internal representation of the real number r, i.e., $r : 1 \to \mathcal{R}$ defined for all A as $r_A(1(A)) = \ulcorner r \urcorner$. Now we have a mathematical machinery sufficient to give the categorical definition of measure. For every A, let P_A be the function that sends $L \in \Delta^{\mathcal{B}}(A) = \mathcal{B}$ to the function $p_A \in \mathcal{R}(A)$ that maps every $\sigma^X_x \in A$ to $\sigma^X_x(L)$. It is easy to see that the family $\{P_A\}_{A \text{ object of } U}$ defines a natural transformation $P : \Delta^{\mathcal{B}} \to \mathcal{R}$.

Definition 4. *The natural transformation $P : \Delta^{\mathcal{B}} \to \mathcal{R}$ is called categorical measure.*

The following lemma justifies the name given to the natural transformation P.

Lemma 1. *For all A, $P(L) = r$ is true in A, i.e. $(P(L) = r)_A(1(A)) = true_A$, iff $(\forall \sigma^X_x \in A)(\sigma^X_x(L) = r)$*

In the internal language of $\mathcal{T}$–*Sets* (for suitable $\mathcal{T}$), possibilistic, probabilistic and imprecise-probabilistic reasoning have a valid and complete representation (see [4–6]).

Note that a proposition in $\mathcal{T}$–*Sets* is a time invariant. In fact $\alpha : 1 \to \Omega$ is true in A (i.e. $\alpha_A(\star) = max^A$) iff α remains true in every B that comes after A (i.e. every B s.t. there exists $f : B \to A$). Moreover $\alpha = true$ iff α is uncertainty invariant, in fact it remains true in every informational state A, i.e. whatever is the (unknown) measure that governs the process. Therefore the *true* formulae of $\mathcal{T}$–*Sets* define time and uncertainty invariant properties. The next task is to interpret proofs into *true* formulae of $\mathcal{T}$. An atomic relation is the relation defined by a clique of an atomic formula or the negation of an atomic formula. A b-relation is Boolean combination of atomic relations. A clique $a_{\bar{x}} \sqsubset |\phi|$ is generic if it has the form $a_{\bar{x}} = \{x : x\rho\bar{x}\}$ for ρ a b-relation. Every generic clique a has a natural interpretation a° into the internal language of $\mathcal{T}$–*Sets*. To see this let me introduce a suitable formula of the internal language of $\mathcal{T}$–*Sets*. If $\sigma_i : \Delta^{|X_i|} \times \Delta^{\mathcal{B}_i} \to \Delta^R$ is defined by $\sigma_i(x_i, L) = \sigma^{X_i}_{x_i}(L)$ then $\phi(x)$ is the formula $(\forall x_{\Delta^{\mathcal{B}_i}})(P(x_{\Delta^{\mathcal{B}_i}}) = \sigma(x_i, x_{\Delta^{\mathcal{B}_i}}))$. If $a_{\bar{x}} \sqsubset P$ then $a_{\bar{x}}$ is made by all observables x s.t. $\sigma^{X_i}_x = \sigma^{X_i}_{\bar{x}}$, therefore it is natural to interpret $a_{\bar{x}}$ into $\phi(\bar{x})$, because $\phi(\bar{x})$ is *true* exactly in the A_x with $\sigma^{X_i}_x = \sigma^{X_i}_{\bar{x}}$. We can interpret 1 in $\mathcal{T}$ with the (*true*) formula $P(\Xi) = 1$. If $a_{\bar{x}} \sqsubset P^\perp$ then $a_{\bar{x}}$ is made by all tests against $\bar{x}$ i.e. it is made by all observables x s.t. $\sigma^{X_i}_x \neq \sigma^{X_i}_{\bar{x}}$ therefore it is natural to interpret $a_{\bar{x}}$ into $\neg\phi(\bar{x})$, because $\neg\phi(\bar{x})$ is *true* exactly in the A_x with $\sigma^{X_i}_x \neq \sigma^{X_i}_{\bar{x}}$. With the same arguments every b-clique can be represented in the corresponding combination of atomic interpretations. It is possible to prove that every cut-free proof in the multiplicative fragment of linear logic is interpreted in a generic clique:

Theorem 1. *If* $\vdash_\pi \phi$ *is a cut free poof then* $(\pi^*)^\circ = true$

References

1. J.-Y. Girard. Linear logic: its syntax and semantics. In J.-Y. Girard, Y. Lafont, and L. Regnier, editors, *Advances in linear logic*. Cambridge University Press, 1995.
2. P.R. Halmos. *Measure Theory*. Van Nostrand, 1950.
3. S. Mac Lane and I. Moerdijk. *Sheaves in geometry and logic: a first introduction to topos theory*. Springer-Verlag, New York, 1992.
4. C. Sossai. U-Sets as probabilistic sets. Technical Report 02/03, ISIB-CNR, Padova, Italy, April 2003.
5. C. Sossai. U-Sets as a possibilistic set theory. *Fuzzy Sets and Systems*, 144(1):193–212, 2004.
6. C. Sossai. Reasoning with uncertainty and context-dependent languages. In *EUSFLAT-LFA 2005*, pages 769–774, Barcelona, Spain, 2005.
7. C. Sossai. Chance as a new semantics for logic. In *IPMU 2006*, pages 1220–1227, Paris, France, 2006.

A Concept of Duality for Multivariate Exchangeable Survival Models

Fabio Spizzichino

Department of Mathematics, University La Sapienza
Piazzale A. Moro 5, 00185 Rome, Italy
fabio.spizzichino@uniroma1.it

Let C be a n-dimensional copula and $\overline{G}:[0,\infty)\rightarrow(0,1]$ a one-dimensional survival function, i.e. a non-increasing function such that

$$\overline{G}(0)=1, \lim_{x\rightarrow\infty}\overline{G}(x)=0.$$

We assume $\overline{G}$ to be continuous and strictly decreasing and consider the multivariate function defined by

$$\overline{F}(x_1,...,x_n):=C\left(\overline{G}(x_1),...,\overline{G}(x_n)\right).$$

$\overline{F}$ can be seen as the *joint survival function* of n lifetimes (i.e., non-negative random variable) $T_1,...,T_n$, i.e. $T_1,...,T_n$ are such that

$$\overline{F}(x_1,...,x_n)=P\{T_1>x_1,...,T_n>x_n\}.$$

We thus say that the pair $\left(C,\overline{G}\right)$ determines a n-dimensional survival model.

$T_1,...,T_n$ are exchangeable, admit C as their *survival copula* and $\overline{G}$ as their *marginal survival function*.

The same model can be alternatively characterized by the pair $[B,\overline{G}]$, where the function $B:[0,1]^n\rightarrow[0,1]$ is the *ageing function*, defined by

$$B(u_1,...,u_n):=\exp\{-\overline{G}^{-1}\left(\overline{F}(-\log u_1,...,-\log u_n)\right)\}$$

The concept of *ageing function* emerges in the field of reliability and, in particular, in the study of multivariate aging properties of a vector of lifetimes; from a mathematical viewpoint, the role of B lies in that it is a function $B:[0,1]^n\rightarrow[0,1]$ (then a fuzzy set of $[0,1]^n$, like copulas) adapt to describe the family of the level curves of the function $\overline{F}$.

Definition, meaning, applications, and different mathematical properties of B have been studied in [1], [2], [3], where the special bivariate case $n=2$ has been considered; most of arguments therein can be immediately extended to the case $N>2$; see also the article [4].

B is component-wise increasing; generally, B is a *semicopula*, but not necessarily a copula, i.e. it satisfies all the properties of a copula, but the rectangular inequality. For general aspects of the concept of semicopula see also [5], [6].

In terms of the pair $[B,\overline{G}]$, $\overline{F}$ can be obtained by writing

$$\overline{F}(x_1,...,x_n)=\overline{G}\left(-\log B\left(e^{-x_1},...,e^{-x_n}\right)\right).$$

In this talk we restrict attention to survival models whose ageing functions are actually copulas. As a main purpose, we introduce a notion of duality for pairs of such models and analyze some basic aspects of this notion.

For a given multivariate survival model $\overline{F}$, denote by $\widehat{C}_{\overline{F}}$, $\overline{G}_{\overline{F}}$ and $B_{\overline{F}}$ the corresponding survival copula, univariate marginal survival function and aging function, respectively.

Two different survival models $\overline{L}$ and $\overline{M}$, with ageing functions belonging to the family of copulas, are *dual* each other if

$$\widehat{C}_{\overline{M}} = B_{\overline{L}}, B_{\overline{M}} = \widehat{C}_{\overline{L}},$$
$$\overline{G}_{\overline{L}}(x) = \exp\{-R(x)\}, \overline{G}_{\overline{M}}(x) = \exp\{-R^{-1}(x)\}.$$

This concept of duality can reveal useful to the purpose of analyzing properties of models with a fixed ageing function B, employing results proved in the literature about copulas.

Among the basic properties that will be presented, we will show a method to construct dual pairs by starting from models with standard exponential marginal.

This piece of research was started years ago in collaboration with Bruno Bassan. In this occasion, I will present recently obtained results along with the basic definitions and properties of duality, that had been worked out at those times and never presented before.

References

1. Bassan Bruno and Spizzichino Fabio (2001). Dependence and multivariate aging: the role of level sets of the survival function. *System and Bayesian Reliability* (Yu Hayakawa, Telba Irony and Min Xie, Eds.), 229–242, World Sci. Publ., River Edge, NJ.
2. Bassan Bruno and Spizzichino Fabio (2003).On some properties of dependence and aging for residual lifetimes in the exchangeable case. *Mathematical and Statistical Methods in Reliability* (Bo H. Lindqvist and Kjell A. Doksum., Eds.), 235–249, World Sci. Publ., River Edge, NJ.
3. Bassan Bruno and Spizzichino Fabio (2005). Relations among univariate aging, bivariate aging and dependence for exchangeable lifetimes. *J. Multivariate Anal.* 93 , no. 2, 313–339.
4. Bassan Bruno and Spizzichino Fabio (2005). Bivariate survival models with Clayton aging functions. *Insurance Math. Econom.* 37, no. 1, 6–12.
5. Durante, Fabrizio and Sempi Carlo (2005). Semicopulae. *Kybernetika*, 41, no. 3, 315–328.
6. Durante, Fabrizio and Sempi Carlo (2005). Copula and semicopula transforms. *Int. J. Math. Math. Sci.*, no. 4, 645–655.

A Note on States on Generalized Residuated Lattices

Esko Turunen and Janne Mertanen*

Tampere University of Technology
P.O. Box 553, 33101 Tampere, Finland
esko.turunen@tut.fi

A mathematical concept *state* dates from 1936 when Birkhoff and von Neumann published their study *The logic of quantum mechanics* [1]. In 1986, Mundici [4] gave fundations for probability theory on MV–algebras and showed the advantages of such approach in quantum logic framework (see also [5, 6]). Precisely, a state on an MV–algebra L is a function $s : L \to [0,1]$ such that $s(\mathbf{0}) = 0$, $s(\mathbf{1}) = 1$ and, for all $a,b \in L$ with $a \odot b = \mathbf{0}$ we have $s(a) + s(b) = s(a \oplus b)$. MV–algebras are known to be special residuated lattices with a proper *additive operation* $\oplus$. MV–algebras fulfill a *double negation law* $x^{**} = x$, too.

Recently, the notion of state has been extended to more general residuated structures L, see [2], [3] and references thereon. Indeed, two different notions have been introduced; a mapping $s : L \to [0,1]$ is

(i) *Riečan state* if $s(\mathbf{1}) = 1$ and

$$s((y^{\sim} \odot x^{\sim})^{-}) = s(x) + s(y) \text{ whenever } y^{-\sim} \odot x^{-\sim} = \mathbf{0},$$

(ii) *Bosbach state* if $s(\mathbf{1}) = 1, s(\mathbf{0}) = 0,$ and, for all $x, y \in L$:

$$s(x) + s(x \to y) = s(y) + s(y \to x),$$
$$s(x) + s(x \Rightarrow y) = s(y) + s(y \Rightarrow x).$$

In [2] it is proved that in any *good generalized* residuated lattice L Bosbach states are Riečan states, while the converse is not always true. Here *generalized* residuated lattice means a residuated lattice whose product operation $\odot$ needs not to be commutative, and the notion *good* has a non–trivial meaning only in non–commutative residuated lattices. Thus, the concept of Riečan state is more general than of Bosbach state. Crucial properties of Riečan states s are that they are (i) isotone, i.e. if $x \leq y$ then $s(x) \leq s(y)$ and (ii) $s(x^{\sim}) = s(x^{-}) = 1 - s(x)$.

After observing that a sort of *orthogonality* of complement elements is required in the definition of Riečan state s and that for any such state it holds that

$$s(x) = s(x^{\sim\sim}) = s(x^{-\sim}) = s(x^{\sim-}) = s(x^{--}) \text{ for any element } x \in L,$$

it is relevant to ask if Riečan states on L are just Riečan states on the generalized MV–subalgebra $MV(L)$ of complement elements $x^{-\sim}$ of L presumed, of course, that such a generalized MV-subalgebra of L exists. We demonstrate that this, indeed, is the case. In [7], we showed that, given a (commutative) residuated lattice L, a subset $MV(L)$ of complement elements x^* of L generates an MV–algebra if, and only if L is semi–divisible. On a semi–divisible residuated lattice L Riečan states and Riečan states on $MV(L)$ are essentially the very same thing. The same holds for Boscach states as far as L is divisible. The aim of this paper is to generalize these results to apply to good generalized residuated lattices (cf. [2]), too.

---

* These results were obtained when the authors visited Academy of Science, Czech Republic, Dept. of Comp. Sciences in Autumn 2006.

References

1. Birkhoff, G. and von Neumann, J.: The logic of quantum mechanics. Ann. Math **37**(1936), 823–843.
2. Ciungu, L.C.: Bosbach and Riečan states on residuated lattices. Manuscript. 15 pages.
3. Dvurečenskij, A. and Rachůnek, J.: Probabilistic Averaging in Bounded Rl–Monoids. Semigroup Forum **72**(2006), 190–206.
4. Mundici, D.: Interpretation of AFC^*–algebras in Lukasiewicz sentential calculus. Journal of Functional Analysis, **65**(1986), 15–63.
5. Mundici, D.: Averaging the Truth-Value in Lukasiewicz Logic, Studia Logica **55**(1995), 113–127.
6. Mundici, D.: Bookmaking over infinite–valued events. Jour.Appr. Reasoning,(2006), to appear.
7. Turunen, E., Mertanen, J.: States on semi–divisible residuated lattices (submitted).

Imprecise Inference Models in Risk Analysis

Lev V. Utkin

St. Petersburg State Forest Technical Academy
Institutski per. 5, 194021 St. Petersburg, Russia
lev.utkin@mail.ru

1 Introduction

Risk can be defined in many ways and has a variety of common meanings. However, the common purpose of risk analysis is to provide decision support for design and operation and, therefore, risk analysis is always part of a decision context. The traditional approach to decision analysis in the framework of expected utility theory calls for single or precise distributions of states of nature. However, we have usually only partial information about probabilities of states of nature. Various tools for sophisticated uncertainty representation generalizing the common ('classical') concept of probability can be found in the literature, including possibility theory, Dempster-Shafer structures, interval-valued probabilities, imprecise probabilities. The corresponding decision making models have been developed in accordance with the different types of the uncertainty representation. In contrast to standard decision theory, these models allow to handle partial information about the stochastic behavior of the states of nature.

The paper consists of two parts. In the first one, we explicitly take into account the construction of the information and consider some decision problems where direct data (precise or interval-valued) on the states are available. Moreover, imprecise models are proposed and studied here by proceeding from certain applications (risk model of insurance). The second part of the paper studies a decision problem in the case of a *non-monotone utility function* when states of nature are described by sets of *continuous probability distributions* restricted by some known lower and upper distributions.

2 Individual risk model of insurance and imprecise models

Let us briefly consider the well-known individual risk model of insurance, which is widely used in applications, especially in life and health insurance. We assume that the portfolio consists of N identical insurance policies for a given period of time t, each insurance premium is c, each policy produces a payment with the claim amount (size) y_i. Then the total premium for the time t is $\Pi(t) = cN$ and the total amount of claims is defined as $R(t)$. The probability that aggregate claims will be less than the premium collected is $P = \Pr\{\Pi(t) \geq R(t)\}$. Suppose the random number k of claims for the time t has a discrete distribution $p(k|w)$ with a parameter (or a set of parameters) w as $P = \sum_{k=0}^{M} p(k|w)$. Here $M = \lceil \Pi(t)/y \rceil = \lceil cN/y \rceil$ is the maximal number of claims which can be paid by the insurer. The next study depends on the distribution function of the number of claims and its parameters w.

If the parameter w is unknown and we have a set of observations (claims), then w would be regarded as a random variable with some probability density $\pi(w|\theta)$ and Bayesian approach could be applied for computing the probability P.

2.1 Imprecise inference models

Imprecise beta-binomial model By binomially distributed numbers of claims with parameter $w = q$, the conjugate distribution π is the beta distribution with parameters $\theta = (a, b)$. Then the posterior distribution is nothing else but the beta-binomial distribution with parameters a and b. Walley [5] proposed the imprecise model which can be defined as the set of all beta-binomial distributions with the fixed parameter s and the set of parameters $0 \leq \alpha \leq 1$. If we replace (a,b) by $(s\alpha, s - s\alpha)$, then the lower and upper bounds for P can be obtained by minimizing and maximizing P over all values α in $[0,1]$. The corresponding expressions can be found in Walley's paper [5].

Imprecise negative binomial model I If the number of claims has the Poisson distribution with the parameter $w = \lambda$, then the conjugate distribution π is the gamma distribution with parameters $\theta = (a,b)$. The probability that aggregate claims will be less than the premium collected for time t is

$$P = \sum_{k=0}^{M} \frac{\Gamma(a+K+k)}{\Gamma(a+K)k!} \cdot \left(\frac{b+T}{b+T+t}\right)^{a+K} \left(\frac{t}{b+T+t}\right)^{k},$$

where K is the observed claim number; T is the observation time; Γ is the gamma function.

Here imprecise probability models for inference in exponential families proposed by Quaeghebeur and de Cooman [1] can be applied. If we replace (a,b) by $(s\alpha, s)$, then the lower and upper bounds for P can be obtained by minimizing and maximizing P over all values α in $[0,\infty)$. Hence $\underline{P} = 0$ and

$$\overline{P} = \sum_{k=0}^{M} \frac{\Gamma(K+k)}{\Gamma(K)k!} \cdot \left(\frac{s+T}{s+T+t}\right)^{K} \left(\frac{t}{s+T+t}\right)^{k}. \tag{1}$$

The main problem of the models proposed by Quaeghebeur and de Cooman [1] is the trivial lower bound $\underline{P}$.

Imprecise negative binomial model II Another model for constructing a set of negative binomial distributions as a reasonable class of priors has been proposed by Coolen and is determined by the set of parameters (a,b) values is within the triangle $(0,0)$, $(s,0)$, $(0,s)$. Here the hyperparameter $s \geq 0$. The interpretation is that all possible prior rates of occurrence of claims are represented, as the prior allows interpretation of a/b as this rate, hence this would include all such rates in $(0,\infty)$. The lower probability $\underline{P}$ is achieved at $(a,b) = (s,0)$ and is determined as

$$\underline{P} = \sum_{k=0}^{M} \frac{\Gamma(s+K+k)}{\Gamma(s+K)k!} \left(\frac{T}{T+t}\right)^{s+K} \left(\frac{t}{T+t}\right)^{k}.$$

The upper probability $\overline{P}$ is achieved at $(a,b) = (0,s)$ and is determined by (1). If $s = 0$, then $\underline{P} = \overline{P}$.

2.2 Interval-valued data

Suppose we have only interval-valued data, for instance, the number of observed claims is from 9 to 10. In other words, we write $K \in [9,10]$ or $K \in [\underline{K}, \overline{K}]$. The similar (but not the same) type of data by using the imprecise Dirichlet model has been investigated by Utkin [3]. However, we can not use the imprecise Dirichlet model here and have to construct another model.

Suppose there are m expert estimates (observations) of K in the form of sets of intervals $K_1, ..., K_m$. Here the i-th interval occurs c_i times. Let $C = c_1 + ... + c_m$. Then the basic probability assignment $m(K_i) = c_i/C$ can be defined for every K_i.

Let $P(K)$ be a function linking the probability P and the values of K. Then we can determine expected probabilities $P^{\#}$ that aggregate claims will be less than the premium collected. Strat [2] proposed expressions for computing bounds for the expected utility by interval-valued data about states of nature. These expressions can be successfully applied to computing the bounds of expected probabilities $P^{\#}$ if we assume that $P(K)$ is the utility function and values of K are states of nature, i.e.,

$$\underline{P}^{\#} = \sum_{i=1}^{m} m(K_i) \cdot \min_{K \in K_i} P(K), \; \overline{P}^{\#} = \sum_{i=1}^{m} m(K_i) \cdot \max_{K \in K_i} P(K).$$

Since $P(K)$ decreases as K increases, then the above expressions can be rewritten as

$$\underline{P}^{\#} = \frac{1}{C} \sum_{i=1}^{m} c_i P(\overline{K}_i), \; \overline{P}^{\#} = \frac{1}{C} \sum_{i=1}^{m} c_i P(\underline{K}_i).$$

In sum, we have obtained the simple expressions for expected lower and upper probabilities that aggregate claims will be less than the premium collected. This model can be simply extended by using the approach proposed by Utkin and Augustin in [4].

It should be noted that the considered imprecise models can be also applied to a scheme of typical warranty contracts, to market research and other applications. Moreover, new imprecise models can be developed by investigating different distributions, for instance, the exponential-gamma model, the Dirichlet-multinomial model, the normal model.

3 Risk analysis with non-monotone utility function

Suppose that information about random variable X, characterizing states of nature, is represented by some lower $\underline{F}$ and upper $\overline{F}$ probability distributions and $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$, $\forall x \in \mathbb{R}$. Then for a function $h(X)$ (utility function), the lower and upper expectations (expected utilities) can be computed as (Choquet integrals)

$$\underline{\mathbb{E}}h = \inf_{\underline{F} \leq F \leq \overline{F}} \int_{\mathbb{R}} h(x) \mathrm{d}F(x), \; \overline{\mathbb{E}}h = \sup_{\underline{F} \leq F \leq \overline{F}} \int_{\mathbb{R}} h(x) \mathrm{d}F(x). \tag{2}$$

If the function h is non-decreasing in $\mathbb{R}$, then there hold

$$\underline{\mathbb{E}}h = \int_{\mathbb{R}} h(x) \mathrm{d}\overline{F}(x), \;\; \overline{\mathbb{E}}h = \int_{\mathbb{R}} h(x) \mathrm{d}\underline{F}(x).$$

The case of the non-increasing function h is similar. It can be seen from the above that the bounds for expectations are completely defined by bounded distributions $\underline{F}$ and $\overline{F}$.

Suppose now that h has one maximum at point x_0, i.e., $h(x)$ is increasing in $(-\infty, x_0]$ and decreasing in $[x_0, \infty)$. In this case, the upper and lower expectations of h are

$$\overline{\mathbb{E}}h = h(x_0) \left[\overline{F}(x_0) - \underline{F}(x_0) \right] + \int_{-\infty}^{x_0} h(x) \mathrm{d}\underline{F}(x) + \int_{x_0}^{\infty} h(x) \mathrm{d}\overline{F}(x),$$

$$\underline{\mathbb{E}}h = \min_{\alpha \in [0,1]} \left[\int_{-\infty}^{\overline{F}^{-1}(\alpha)} h(x) \mathrm{d}\overline{F}(x) + \int_{\underline{F}^{-1}(\alpha)}^{\infty} h(x) \mathrm{d}\underline{F}(x) \right]. \tag{3}$$

Moreover, the minimum over $\alpha \in [0,1]$ in (3) is achieved at a point which is one of the solutions to the equation $h\left(\overline{F}^{-1}(\alpha)\right) = h\left(\underline{F}^{-1}(\alpha)\right)$.

Now we consider a general form of the function h, i.e., the function has alternate points of the local maximum at a_i and minimum at b_{i-1}, $i = 1, 2, \ldots$, such that $b_0 < a_1 < b_1 < a_2 < b_2 < \ldots$. The solution to the optimization problem for computing $\overline{\mathbb{E}}h$ is the function $F(x) = F_i(x)$, $x \in (b_{i-1}, b_i)$, with jumps at points b_i. The size of the i-th jump is $\min\left(\overline{F}(b_i), \alpha_{i+1}\right) - \max\left(\underline{F}(b_i), \alpha_i\right)$. Here

$$F_i(x) = \begin{cases} \overline{F}(x), & x < a' \\ \alpha, & a' \leq x \leq a'' \\ \underline{F}(x), & a'' < x \end{cases},$$

where α is the root of the equation $h\left(\max\left(\overline{F}^{-1}(\alpha), b_{i-1}\right)\right) = h\left(\min\left(\underline{F}^{-1}(\alpha), b_i\right)\right)$ in interval $\left[\underline{F}(a_i), \overline{F}(a_i)\right]$, $a' = \max\left(\overline{F}^{-1}(\alpha), b_{i-1}\right)$, $a'' = \min\left(\underline{F}^{-1}(\alpha), b_i\right)$.

The proposed approach is a way for avoiding computationally difficult procedures for solving (2) by means of approximate linear programming.

References

1. E. Quaeghebeur and G. de Cooman. Imprecise probability models for inference in exponential families. In J.-M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *Proc. of the 4rd Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'05*, Pittsburgh, Pennsylvania, July 2005. Carnegie Mellon University.
2. T.M. Strat. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4(5):391–418, 1990.
3. L.V. Utkin. Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model. *Fuzzy Sets and Systems*, 154(3):413–431, 2005.
4. L.V. Utkin and Th. Augustin. Decision making under imperfect measurement using the imprecise Dirichlet model. In T. Seidenfeld F.G. Cozman, R. Nau, editor, *Proc. of the 4th Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'05*, pages 359–368, Pittsburgh, USA, July 2005. Carnegie Mellon University, SIPTA.
5. P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996. with discussion.

Subalgebras of the Algebra of Truth Values of Type-2 Sets and Their Automorphisms

Carol Walker and Elbert Walker

Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003
{hardy|elbert}@nmsu.edu

Type-2 fuzzy sets—that is, fuzzy sets with fuzzy sets as truth values were introduced by Zadeh [4], extending the notion of ordinary fuzzy sets. In [1], there is a treatment of the mathematical basics of type-2 fuzzy sets, that is, of the algebra of these truth values. In [2], a study was begun of automorphisms of the algebra of truth values of type-2 fuzzy sets. Several significant problems left unresolved are now settled, the most basic being that the automorphism group of the algebra of fuzzy truth values is isomorphic in a natural way to the product of two copies of the automorphism group of the unit interval with its usual ordering. This theorem has several corollaries concerning characteristic subalgebras and their automorphism groups. This paper is about those things: the automorphisms of the algebra of fuzzy truth values, its subalgebras, and their automorphisms.

For any mathematical object, its group of symmetries, which for algebraic structures are called automorphisms, is an object of interest. The study of subalgebras of the algebra of truth values of type-2 fuzzy sets is relevant because each subalgebra is the basis of a fuzzy theory, where a fuzzy set in this theory is a mapping of a universal set into this subalgebra. Two well known subalgebras are (isomorphic copies of) the algebras of truth values of type-1 fuzzy sets, and of the truth values of interval-valued fuzzy sets. But there are many others of interest, mathematically, and possibly for applications.

The subalgebras considered are typically characteristic. That is, automorphisms of the algebra of truth values induce automorphisms of these subalgebras. Characteristic subalgebras are of special interest because they are "canonical". If an algebra is characteristic, then there is no subalgebra isomorphic to it sitting in the containing algebra in the same way. They are quite special as subalgebras. That the algebras of truth values of type-1 fuzzy sets and the truth value algebra of interval-valued fuzzy sets are characteristic subalgebras is testimony to the "correctness" of Zadeh's generalization.

Some subalgebras of the algebra of truth values of type-2 fuzzy sets may be viewed much more simply than as such subalgebras. Specifically, the basic operations of the algebra of truth values are convolutions of functions, and some subalgebras may be viewed as algebras with much simpler operations, both conceptually and computationally. This is true, for example, for the subalgebra of closed intervals, as pointed out in [1]. Another such subalgebra is the subalgebra of points: the subalgebra of functions whose support is a single point. This subalgebra generalizes in a particular way the truth value algebra of type-1 fuzzy sets and seems a reasonable candidate for applications. Still another is the subalgebra of those functions whose support is a closed interval and which are constant on that interval. This algebra generalizes the truth value algebra of interval-valued fuzzy sets in the same spirit as points generalize that of type-1 fuzzy sets. These may be viewed as algebras whose basic operations are particularly simple, avoiding complicated computations with convolutions.

1 The Algebra of Fuzzy Truth Values

The algebra of truth values for fuzzy sets of type-2 is the set of all mappings of $[0,1]$ into $[0,1]$ with operations certain convolutions of operations on $[0,1]$, as follows.

Definition 1. *On* $[0,1]^{[0,1]}$*, let*

$$(f \sqcup g)(x) = \bigvee_{y \vee z = x} (f(y) \wedge g(z)) \tag{1}$$

$$(f \sqcap g)(x) = \bigvee_{y \wedge z = x} (f(y) \wedge g(z))$$

$$f^*(x) = \bigvee_{y\prime = x} f(y) = f(x')$$

$$\bar{1}(x) = \begin{cases} 1 \textit{ if } x = 1 \\ 0 \textit{ if } x \neq 1 \end{cases}$$

$$\bar{0}(x) = \begin{cases} 1 \textit{ if } x = 0 \\ 0 \textit{ if } x \neq 0 \end{cases}$$

The algebra $\mathbf{M} = \left([0,1]^{[0,1]}, \sqcup, \sqcap, ^*, \bar{0}, \bar{1}\right)$ is the basic algebra of truth values for type-2 fuzzy sets, and is analogous to the algebra $([0,1], \vee, \wedge, ', 0, 1)$, which is basic for type-1 or ordinary fuzzy set theory.

Determining the properties of the algebra $\mathbf{M}$ is helped by introducing the following auxiliary operations.

Definition 2. *For* $f \in \mathbf{M}$*, let* f^L *and* f^R *be the elements of* $\mathbf{M}$ *defined by*

$$f^L(x) = \vee_{y \leq x} f(y) \tag{2}$$

$$f^R(x) = \vee_{y \geq x} f(y)$$

The point of this definition is that the operations $\sqcup$ and $\sqcap$ in $\mathbf{M}$ can be expressed in terms of the pointwise max and min of functions, as follows.

Theorem 1. *The following hold for all* $f, g \in \mathbf{M}$.

$$\begin{aligned} f \sqcup g &= \left(f \wedge g^L\right) \vee \left(f^L \wedge g\right) \\ &= (f \vee g) \wedge \left(f^L \wedge g^L\right) \end{aligned} \tag{3}$$

$$\begin{aligned} f \sqcap g &= \left(f \wedge g^R\right) \vee \left(f^R \wedge g\right) \\ &= (f \vee g) \wedge \left(f^R \wedge g^R\right) \end{aligned} \tag{4}$$

Using these auxiliary operations, it is fairly routine to verify the following properties of the algebra $\mathbf{M}$. The details may be found in [1].

Corollary 1. *Let* $f,\ g,\ h \in \mathbf{M}$*. The basic properties of* $\mathbb{M}$ *follow.*

1. $f \sqcup f = f$; $f \sqcap f = f$
2. $f \sqcup g = g \sqcup f$; $f \sqcap g = g \sqcap f$
3. $\bar{1} \sqcap f = f$; $\bar{0} \sqcup f = f$

4. $f \sqcup (g \sqcup h) = (f \sqcup g) \sqcup h$; $f \sqcap (g \sqcap h) = (f \sqcap g) \sqcap h$
5. $f \sqcup (f \sqcap g) = f \sqcap (f \sqcup g)$
6. $f^{**} = f$
7. $(f \sqcup g)^* = f^* \sqcap g^*$; $(f \sqcap g)^* = f^* \sqcup g^*$

It is not known whether or not every equation satisfied by **M** is a consequence of these. As far as we know, the variety generated by **M** has not been studied.

In [1], we studied the algebra **M** and some of its subalgebras, and in [2] and [3] their automorphisms. In our study of automorphisms, we limit ourselves initially to the algebra

$$\mathbb{M} = ([0,1]^{[0,1]}, \sqcup, \sqcap, \bar{0}, \bar{1}),$$

that is, the algebra **M** without its negation *. This allows more automorphisms, avoids certain technicalities, and it turns out that the results can be specialized to **M.**

2 Automorphisms of $\mathbb{M}$

For an automorphism α of $\mathbb{I} = ([0,1], \vee, \wedge, 0, 1)$, α_L and α_R defined by $\alpha_L(f) = \alpha f$ and $\alpha_R(f) = f\alpha$ are automorphisms of $\mathbb{M}$. These automorphisms satisfy

1. $(\alpha\beta)_L = \alpha_L\beta_L$
2. $(\alpha\beta)_R = \beta_R\alpha_R$
3. $\alpha_L\beta_R = \beta_R\alpha_L$

The principal result about automorphisms is that every automorphism of $\mathbb{M}$ is of the form $\alpha_L\beta_R$, and uniquely so. Thus $Aut(\mathbb{M}) \approx Aut(\mathbb{I}) \times Aut(\mathbb{I})$. This has many corollaries. For example, the subalgebras mentioned earlier are all characteristic. One subalgebra of special interest is the subalgebra of normal convex functions. It is a maximal lattice among subalgebras of $\mathbb{M}$, is a characteristic subalgebra of $\mathbb{M}$, and is a complete lattice. There are many other results in the same vein.

A basic tool in our investigation is the determination of the irreducible elements of $\mathbb{M}$. An element f is *join irreducible* if $f = g \sqcup h$ implies $f = g$ or $f = h$. *Meet irreducible* is defined similarly, and an element is *irreducible* if it is both join and meet irreducible. The irreducible elements are determined for $\mathbb{M}$ and for various of its subalgebras, enabling the determination of their automorphism groups. However, many questions remain, both concerning automorphisms of subalgebras of $\mathbb{M}$ and other algebraic aspects of $\mathbb{M}$ and its subalgebras. These will be elaborated on.

References

1. Walker, C. and Walker, E.: The Algebra of Fuzzy Truth Values, Fuzzy Sets and Systems **149** (2005) 309-347.
2. Walker, C. and Walker, E.: Automorphisms of the Algebra of Fuzzy Truth Values, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, accepted.
3. Walker, C. and Walker, E.: Automorphisms of the Algebra of Fuzzy Truth Values II, preprint.
4. Zadeh, L.: The Concept of a Linguistic Variable and Its Application to Approximate Reasoning, Inform. Sci. **8** (1975) 199-249.

Uncertainty Measures — Problems Concerning Additivity

Siegfried Weber

Institut für Mathematik, FB 08, Universität Mainz, Germany
sweber@mathematik.uni-mainz.de

We consider an uncertainty measure m on a lattice L with an additional residual structure in the sense of [1-4] as an isotonic mapping from L to [0,1] with m(0)=0, m(1)=1. There are two natural properties which m can have, namely the valuation based on the lattice operatons, and the additivity based on the operations derived from the residual structure. If L is an MV-algebra, the meaning of additivity is clear and well known, as several authors had found independently many years ago. Furthermore, a simple exercise shows that for MV-algebras additivity implies valuation.

Problems arise if we consider a residual structure on L poorer than an MV-algebra. Dropping only the divisibility, we are led to a Girard-algebra, so called in [2-4]. This structure has been found as the natural one for the set $\tilde{L}$ of pairs (a,b) of "events" in L with a≤b, which can be used to define "conditional events". Because in [2,4] it was proved that any Girard-algebra L has a unique canonical Girard-algebra extension $\tilde{L}$. Furthermore, there was shown that $\tilde{L}$ is an MV-algebra iff L is a Boolean algebra, and that an additive measure m on a Boolean algebra L has a unique additive extension $\tilde{m}$ on the canonical MV-algebra extension $\tilde{L}$. But if we start with an additive measure m on an MV-algebra L, it is not clear what "additivity" of an extension $\tilde{m}$ on the canonical Girard-algebra $\tilde{L}$ means, because the "classical additivity" leads to contradictions.

It seems that the additivity of a measure is strongly connected with the divisibility of the underlying lattice. Therefore, in [2,4] **additivity of a measure** on a Girard-algebra was proposed to be defined by the "classical additivity" not for all disjoint pairs, but only for pairs which have the divisibility property. These pairs will be called, for short, **admissible pairs**. Now it has sense to reask the question whether an additive measure m on an MV-algebra L has an extension $\tilde{m}$ on the canonical Girard-algebra $\tilde{L}$ which is additive and resp. or valuation. The answer to both questions in general is negative in the sense that there are examples where neither an additive nor a valuation extension exist, and there are examples where both exist but they are different. As a positive answer, there can be given a non trivial example for a unique extension which is both additive and valuation. More general, denoting by a' the residual complement of a , we can proof the following (for me surprising) result:

If m is additive on an MV-algebra L and $\tilde{m}$ is both additive and valuation on $\tilde{L}$, then it follows (*): $\tilde{m}(a,b) = [m(a)+m(b)]/2 + [m(b \wedge b')-m(a \wedge a')]/6$.

As a corollary we obtain the above mentioned result from [2,4]:
For a Boolean algebra L, (*) reduces to $\tilde{m}(a,b) = [m(a)+m(b)]/2$.

Furthermore, not all measures of the form (*) really are additive and valuation, but we can caracterize all measures of the form (*) by both, the additivity only for two special types of admissible pairs, and the valuation only for one special type of non admissible pairs.

Returning to the possible situations where neither additive nor valuation measure extensions exist or where both exist but they are different, we can resp. shall look for another weaker notion of additivity. For this reason we propose to define the **weak additivity of a measure** on a Girard-algebra by the **additivity on all sub-MV-algebras**. Really, this is a weaker property because we have to check "classical additivity" not for all admissible pairs, but only for those in the same sub-MV-algebra. In this sense, in all examples we could found several or unique weak additive extensions.

References

[1] S. Weber, *Conditioning on MV-algebras and additive measures, I*, Fuzzy Sets and Systems 92 (1997), 241-250.

[2] U. Höhle and S. Weber, *Uncertainty measures, realizations and entropies,* in: Random Sets: Theory and Applications (eds. J. Goutsias, R.P.S. Mahler, H.T. Nguyen), 259-295 (Springer, 1997)

[3] S. Weber, *Conditioning on MV-algebras and additive measures, further results*, in: Fuzzy Sets, Logics and Reasoning about Knowledge (eds. D. Dubois, H. Prade, E.P. Klement), 175-199 (Kluwer, 1999)

[4] U. Höhle and S. Weber, *On Conditioning Operators*, in: Mathematics of Fuzzy Sets – Logic, Topology and Measure Theory (eds. U. Höhle and S. Rodabaugh), 653-673 (Kluwer, 1999)

Asymptotic Tests for the Aumann-Expectation of Fuzzy Random Variables

Andreas Wünsche

TU Bergakademie Freiberg,
Institut für Stochastik,
Prüferstr. 9, 09599 Freiberg, Germany
wuensche@math.tu-freiberg.de

Abstract. We consider the two-sided asymptotic test for the Aumann-expectation of fuzzy random variables, proposed by Körner [3]. Unfortunately this test is not easy to use. Starting from the question, what is the power of the test, we get theorem 1. With this theorem we can propose a more applicable test for interval hypotheses.

1 Introduction

Vague outcomes of an experiment can be described by fuzzy sets. Following [10], a fuzzy set $\widetilde{A}$ on $\mathbb{R}^d$ is identified by its membership function $\mu_{\widetilde{A}} : \mathbb{R}^d \to [0,1]$. The crisp set$\widetilde{A}_\alpha := \left\{x \in \mathbb{R}^d : \mu_{\widetilde{A}}(x) \geq \alpha\right\}$, $0 < \alpha \leq 1$ is called the α-cut of $\widetilde{A}$. For $\alpha = 0$ define: $\widetilde{A}_0 := closure\left\{x \in \mathbb{R}^d : \mu_{\widetilde{A}}(x) > 0\right\}$, which is called the support of $\widetilde{A}$ (supp$\widetilde{A} = \widetilde{A}_0$). A fuzzy set $\widetilde{A}$ is called convex and compact if all α-cuts $\widetilde{A}_\alpha$ have this property. $\widetilde{A}$ is called normal if $\widetilde{A}_1 \neq \emptyset$. The set of all normal compact convex fuzzy sets on $\mathbb{R}^d$ with bounded support is denoted by $\mathcal{F}_c^d$. For any compact convex set $A \subset \mathbb{R}^d$ the support function s_A is defined as$s_A(u) := \sup_{y \in A} \langle u, y \rangle$; $u \in \mathbb{S}^{d-1}$, where $\langle .,. \rangle$ is the scalar product in $\mathbb{R}^d$ and $\mathbb{S}^{d-1}$ the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$. Note that for convex and compact $A \subset \mathbb{R}^d$ the correspondence between A and s_A is one to one. A fuzzy set $\widetilde{A} \in \mathcal{F}_c^d$ can be characterized α-cut-wise by its support function:

$$s_{\widetilde{A}}(u,\alpha) := s_{\widetilde{A}_\alpha}(u) \; ; \; \alpha \in [0,1], \; u \in \mathbb{S}^{d-1}.$$

Via support function s, $\mathcal{F}_c^d$ can be embedded into a space of functions on $\mathbb{S}^{d-1} \times [0,1]$ and we can define a metric in $\mathcal{F}_c^d$ using e.g. a special L^2-metric in $L^2(\mathbb{S}^{d-1} \times [0,1])$, i.e.

$$\delta_2(\widetilde{A},\widetilde{B}) = \left(d \int_0^1 \int_{\mathbb{S}^{d-1}} (s_{\widetilde{A}}(u,\alpha) - s_{\widetilde{B}}(u,\alpha))^2 \nu(du) d\alpha\right)^{\frac{1}{2}},$$

where ν is the normalized Lebesgue measure on $\mathbb{S}^{d-1}$. We define further

$$\left\langle \widetilde{A},\widetilde{B} \right\rangle := d \int_0^1 \int_{\mathbb{S}^{d-1}} (s_{\widetilde{A}}(u,\alpha) s_{\widetilde{B}}(u,\alpha)) \nu(du) d\alpha \text{ and } ||\widetilde{A}||_2^2 := \left\langle \widetilde{A},\widetilde{A} \right\rangle.$$

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. Then $\widetilde{Y} : \Omega \to \mathcal{F}_c^d$ is called a fuzzy random variable (frv) on $\mathbb{R}^d$ if for any $\alpha \in [0,1]$ the α-cut $\widetilde{Y}_\alpha$ is a convex compact random set (e.g. in the sense of [5]). This is the Puri/Ralescu-approach to frv's ([7]), the essential advantage of which is the embedding of the concept of a frv into the well-developed concept of random sets. For a unified approach see also [4].

The Aumann-expectation of a frv $\widetilde{Y}$ is defined as the fuzzy set $\mathbf{E}^{(A)}\widetilde{Y} \in \mathcal{F}_c^d$ with

$$\forall \alpha \in [0,1] : \quad \left(\mathbf{E}^{(A)} Y\right)_\alpha = \mathbf{E}^{(A)} \widetilde{Y}_\alpha$$

where $\mathbf{E}^{(\mathbf{A})}\widetilde{Y}_\alpha$ is the Aumann-expectation (see [1]) of the random set $\widetilde{Y}_\alpha$ defined by

$$\mathbf{E}^{(A)}\widetilde{Y}_\alpha = \left\{\mathbf{E}X : X(\omega) \in \widetilde{Y}_\alpha \; P-\text{a.e. and } X \in L^1(\Omega, \mathfrak{F}, P)\right\}.$$

$\mathbf{E}^{(A)}\widetilde{Y}_\alpha$ is the set of all (usual) expectations of random "selectors" X which P-a.e. lie in $\widetilde{Y}_\alpha$.

2 The tests proposed by Körner

We consider now a iid. sample of frv $\widetilde{X}_1, \widetilde{X}_2, \ldots$. For the Aumann-expectation of this frv we want to test the following two-sided hypotheses.

$$\begin{array}{ccc} H_0 : \mathbf{E}\widetilde{X} = \widetilde{\mu}_0 & \text{against} & H_0 : \mathbf{E}\widetilde{X} \neq \widetilde{\mu}_0, \\ & \Longleftrightarrow & \\ H_0 : \delta_2^2(\mathbf{E}\widetilde{X}, \widetilde{\mu}_0) = 0 & \text{against} & H_0 : \delta_2^2(\mathbf{E}\widetilde{X}, \widetilde{\mu}_0) \neq 0, \end{array}$$

Körner[3] proposed to use the statistic:

$$T = n\delta_2^2(\overline{\widetilde{X}}, \widetilde{\mu}_0).$$

He proofs that T under $H_0 : \mathbf{E}\widetilde{X} = \widetilde{\mu}_0$ is asymptotic distributed like

$$\sum_{i=1}^{\infty} \lambda_i Z_i^2 \quad \text{where} \quad Z_i \sim \mathcal{N}(0,1) \text{ iid.},$$

where $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of the covariance operator of $\widetilde{X}_i$. So we have to reject H_0 if $T > q_{1-\alpha}$, where $q_{1-\alpha}$ is the quantile of the distribution of $\sum_{i=1}^{\infty} \lambda_i Z_i^2$.

For the practical use of this result one has followings problems:

(i) The covariance operator of $\widetilde{X}_i$ is unknown in generally.
(ii) The numerical calculation of the quantile of the distribution of $\sum_{i=1}^{\infty} \lambda_i Z_i^2$ is not easy.

To overcome the problem (i) one can estimate the eigenvalues $(\hat{\lambda}_i)$ of the covariance operator from the sample $\widetilde{X}_1, .., \widetilde{X}_n$. If we do this, an open question is, under which assumption the asymptotic distribution of $\sum_{i=1}^{\infty} \widehat{\lambda}_i Z_i^2$ is equal to the distribution of $\sum_{i=1}^{\infty} \lambda_i Z_i^2$.

For numerical calculation of the quantile one can use the way which is described in Rice [8]. An other more easy way, to overcome problem (ii), is the simulation of the distribution and then take the quantile from the simulation for the test decision.

The way we will propose in this paper is to soften H_0. I.e. we will not test whether $\mathbf{E}\widetilde{X}$ is equal $\widetilde{\mu}_0$ or not, but only whether it is near $\widetilde{\mu}_0$ or not.

Before we do this we present the main mathematical result of this paper.

3 The main result

If we look at the question, what is the distribution of the statistic $T = n\delta_2^2(\overline{\widetilde{X}},\widetilde{\mu}_0)$ in the case that $H_1 : \mathbf{E}\widetilde{X} \neq \mu_0$ holds, we obtain, which generalizes a result by Montenegro, Colubi, Casals and Gil [6] where only frv's are considered with finite number of possible values.

Theorem 1. *Let $\widetilde{X}_1,\widetilde{X}_2$.. be iid. Aumann-integrable Fuzzy Random variable with $||\widetilde{X}_1||_2^2 < \infty$ and $\widetilde{\mu} := \mathbf{E}\widetilde{X}_1$, then it holds:*

$$\sqrt{n}\left(\delta_2^2(\overline{\widetilde{X}}_n,\widetilde{\mu}_0) - \delta_2^2(\widetilde{\mu},\widetilde{\mu}_0)\right) \xrightarrow{d} \mathcal{N}(0,4\sigma^2),$$

with

$$\sigma^2 = \mathbf{E}\left(\left\langle \widetilde{X}_1 - \widetilde{\mu}, \widetilde{\mu} - \widetilde{\mu}_0 \right\rangle\right).$$

An first conclusion of this theorem is that the test of Körner is consistent. The same result is obtained by Jimenez-Gamero, Pino-Mejýas and Rojas-Medar [2] in their Corollary 3.2.

Corollary 1. *Under the conditions of theorem 1, for $\widetilde{\mu} \neq \widetilde{\mu}_0$ it holds:*

$$\lim_{n\to\infty} P_{\widetilde{\mu}}(T > q_{1-\alpha}) = 1.$$

4 Tests for interval hypotheses

With the theorem 1 we present a proposal for an asymptotic test for the following interval hypotheses:

$$H_0 : \delta_2^2(\mathbf{E}\widetilde{X},\widetilde{\mu}_0) \leq a \qquad \text{against} \qquad H_0 : \delta_2^2(\mathbf{E}\widetilde{X},\widetilde{\mu}_0) > a.$$

With the statistic

$$T = \frac{\sqrt{n}(\delta_2^2(\overline{\widetilde{X}},\widetilde{\mu}_0) - a)}{2\sigma},$$

reject H_0 if $T \geq z_{1-\alpha}$.

Here $z_{1-\alpha}$ is the quantile of the standard normal distribution, which is much more easier to calculate as $q_{1-\alpha}$ from above. The only problem of this test is, that σ is unknown in generally. So we have to replace σ by a consistent estimator. Also here this estimation is much more easier as the estimation of the eigenvalues of the covariance operator above.

5 Conclusion Remarks

For the proof of theorem 1 and more details see Wünsche [9]. Also a more detailed article is in preparation for submission in the Journal of Statistical Planning and Inference.

References

1. Robert J. Aumann: Integrals of Set-Valued Functions. Journal of Mathematical Analysis and Applications, 12, 1-12, (1965)
2. M.D. Jimenez-Gamero,R. Pino-Mejýas,M.A. Rojas-Medar: A bootstrap test for the expectation of fuzzy random variables, Preprint submitted to Comput. Statist. Data Anal., (2004)
3. R. Körner: An asymptotic α-test for the expectation of random fuzzy variables. Journal of Statistical Planning and Inference, 83, No. 2, 331-346 (2000)
4. Volker Krätschmer: A unified approach to fuzzy random variables. Fuzzy Sets and Systems, 123, 1–9, (2001)
5. G. Matheron: Random sets and integral geometry. John Wiley & Sons Inc, John Wiley & Sons Inc (1975)
6. M. Montenegro, A. Colubi, M.R. Casals and M.A. Gil: Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. Metrika, 59, 3149, (2004)
7. Madan L. Puri and Dan. A. Ralescu: Fuzzy random variables. Journal of Mathematical Analysis and Applications, 114, 409–422 (1986)
8. S. O. Rice: Distribution of quadratic forms in normal random variables evaluation by numerical integration. SIAM J. Sci. Stat. Comput., 1, No. 4, 438-448 (1980)
9. A. Wünsche: Statistische Tests bei Unschärfe. PhD Thesis, TU Bergakademie Freiberg, (2005)
10. L. A. Zadeh: Fuzzy Sets. Information and Control, 8 338–353,(1965)

Fuzzy Logic Laboratorium Linz-Hagenberg
Dept. of Knowledge-Based Mathematical Systems
Johannes Kepler Universität
A-4040 Linz, Austria
Tel.: +43 732 2468 9194
Fax: +43 732 2468 1351
E-Mail: info@flll.jku.at
WWW: http://www.flll.jku.at/

Software Competence Center Hagenberg
Hauptstrasse 99
A-4232 Hagenberg, Austria

Tel.: +43 7236 3343 800
Fax: +43 7236 3343 888
E-Mail: sekr@scch.at
WWW: http://www.scch.at/

FUZZY.LOGIC.LAB.LINZ

scch
software competence center
hagenberg

EUSFLAT

EUROPEAN SOCIETY
FOR FUZZY LOGIC
AND TECHNOLOGY