

Towards objective Bayesian foundations with fuzzy events

Pedro Terán*

Departamento de Estadística e I.O. y D.M. Universidad de Oviedo (Spain)

Abstract

We present an axiomatic approach to conditional measures of fuzzy events which is inspired by the Dupré–Tipler argument for Bayesianism. Unlike there or in the older and related Cox argument, although we are able to derive additive probabilities other non-additive solutions exist. Our motivation is to show that Bayesian foundational arguments are based on Boolean logic but they do not require it and may be based on fuzzy logic instead.

Keywords: Conditional possibility, conditional probability, Cox theorem, foundations of statistics, fuzzy event, triangular norm.

1. Introduction

There is a large literature trying to justify Bayesianism, mostly because it is far from self-evident that degrees of belief, certainty, plausibility, or rational expectations must satisfy the same mathematical rules as frequencies. Unfortunately, this necessity has led some to interpret the perceived reasonability of Bayesian methods as grounds for the *strong adequacy thesis* that the theory of probability measures on σ -algebras (or finitely additive probability measures on Boolean algebras) is the unique mathematical tool valid to handle any one among a number of vaguely interlaced tasks concerning rational reasoning under uncertainty.

What are objective logical Bayesians and their Cox argument? The strong adequacy thesis is particularly upheld by a sub-community of Bayesians which has been gaining traction for some years since the posthumous publication of Jaynes' book [7]. They are considered objective Bayesians in that their choice of priors is not based on personal belief but on objective methods like maximum entropy or symmetry arguments. Further, their prototypical task for probability is *plausible reasoning* (a topic going back to [14]), i.e. they spouse a logical view of it.

Their standard argument for the necessity of the probability calculus is based on the Cox theorem [2, 17]. It starts with a short list of 'commonsense' requirements for reasoning with propositions whose truth is unknown, which are translated mathematically as a list of functional equations. Some mathematical assumptions are added in order to solve the

equations and show that all solutions are bijective transforms of finitely additive probability measures.

The argument from the Cox theorem has repeatedly been claimed to rule out, in particular, fuzzy logic. To name just an instance, according to Skilling [15, p. 30] '*Cox ... showed that no other calculus is admissible ... insofar as the various rules of fuzzy logic ... diverge from probability calculus, logical error enters*'. It is unclear how this position could be defensible: the Cox argument already takes Boolean logic as true, whence identifying the weaker fuzzy logics with 'logical error' seems self-contradictory.

Why is it interesting to try and replace Boolean logic by fuzzy logic in the Cox argument? Zadeh said that probability theory should be based on fuzzy logic, but this claim is not easily parsed by Bayesians. By developing a Cox argument compatible with fuzzy logic, it would be clearer to (objective logical) Bayesians that fuzzy logic does not go against their foundational thinking. Since they believe those arguments to be forceful, they are not likely to just take Kosko's dictum that '*Cox's argument is limited by its forthright bivalent-logic framework ... Outside the Bayesian/frequentist debate ... it is hard to imagine citing Cox's theorem as a constraint on anything*' [10, p. 140].

That is similar to what actually happened recently with papers showing that the Dutch book argument for Bayesianism does not necessitate Boolean logic, e.g. [18, 12, 11], whence the Boolean Dutch book argument *just cannot* be an argument against other logics. The following quotation from Hunter [6, p. 26–27] is representative of how those past attacks had been: '*The max-min rules ... in fuzzy logic ... [are] rivals of standard probability theory ... Attempts to replace probability theory wrongly ignore the coherence arguments for the axioms of probability theory given by de Finetti and others ... These arguments show that anyone whose degrees of belief violate the probability axioms could have a "Dutch book" made against him... Fuzzy logic and other rivals of probability face a formidable challenge*'.

It may be that, if a fuzzy Cox argument is feasible, by weakening some of the logical assumptions more solutions will appear other than probability measures. It is worth checking whether this is the case, as it would indicate that premises considered self-evidently convincing by many do not actually

*teranpedro@uniovi.es

support by themselves the strong adequacy thesis.

What is the state of the art in non-Boolean Cox arguments? There is situation of confusion. Knuth has written a number of papers, e.g. [8, 9, 4], extending the Cox theorem to lattices instead of Boolean algebras. However, there are serious mathematical problems in them, as the key step in his papers is the evidently false claim that every real-valued associative function on a lattice is isomorphic to the sum (as an example, just take the maximum in $[0, 1]$).

Holik *et al.* [5] have tried to adapt the Cox argument to quantum logic (which would include Łukasiewicz fuzzy logic) but they make the same mistake of taking it that associative functions are isomorphic to the sum.

Therefore, a study of the feasibility of a fuzzy-based Cox theorem has to start from scratch.

Why is this problem hard and may take time? The proofs of variants of the Cox theorem, in the cases which are mathematically correct, are very long and formulas from Boolean logic which are not valid in one or other fuzzy logic may be used many times.

The most careful proof from explicitly laid out axioms seems to be Van Horn's [17]. In it, for example Axiom R2.(3) says that, if a state of information X grows by consecutively learning two propositions A and C , the new state of information is the same as if their conjunction had been learned. Without idempotency, learning A (e.g. 'blue') and then learning A (i.e. learning nothing new) is not the same as learning $A \cap A$ (e.g. 'very blue').

Thus, problems arise even before the first line of the proof (which then goes up to 'Theorem 14'). Definitely, some stamina is needed for the task.

What is the alternative route taken in this paper? Recently, Dupré and Tipler [3] claimed to have found 'a proof of the problematic Cox theorem which is very simple, in fact trivial, as it should be' [3, p. 599]. What they in fact do (in my opinion, if not theirs) is to present a theorem bearing some undeniable family resemblance to Cox's but which proceeds from axioms of a different nature.

They start with a commutative unital \mathbb{R} -algebra endowed with a partial order, and set requirements for a functional to be a good assessment of the *plausible value* of an element x of the algebra. The plausible value $PV(x|i)$ is conditional on some element i , which is taken to be idempotent (intuitively, x plays the role of a random variable, i the indicator function of an event, the product operation the intersection, and PV the expectation functional). They conclude that

$$PV(i' \cdot i''|i) = PV(i'|i) \cdot PV(i''|i' \cdot i),$$

i.e. the product rule, and

$$xy = 0 \Rightarrow PV(x + y|i) = PV(x|i) + PV(y|i),$$

i.e. the sum rule.

We will adapt the Dupré–Tipler construction to structures of fuzzy sets. Since one cannot rely on

idempotency, which plays a key role in [3], and must have general fuzzy events at both sides of the conditioning bar, and lacks a significant part of the algebraic structure, some leeway is needed. But, even with those modifications, the essence of the Dupré–Tipler argument works well to establish the 'product rule' (better called the 'combination rule' here, see Theorem 3.1 below). We will not even need two of their five axioms, and other two can be weakened in obtaining some of the results, including the combination rule¹.

In the realm of fuzzy events, the combination rule is not strong enough to imply additivity, and the usage of t-norms other than the product is justified. Thus, other solutions appear, specially under the weakened axioms. In particular, one can justify conditional possibilities by resorting to 'kosher' Bayesian argumentation that also produces conditional probabilities.

2. The axioms

For the sake of comparison, Dupré and Tipler's axioms are as follows. Note that these axioms *will not* be used in the sequel².

Axiom DT1. Let \mathbb{E} be a commutative unital algebra over \mathbb{R} , and I its Boolean algebra of idempotent elements. \mathbb{E} is equipped with a partial order \leq such that, for any $i, i' \in I$, we have $i' \leq i$ if and only if $i' = i \cdot i''$ for some $i'' \in I$. Let \mathbb{A}_+ be the set of all non-zero elements of a fixed Boolean subalgebra \mathbb{A} of I which contains the unit 1 of \mathbb{E} . There exists a function $PV : \mathbb{E} \times \mathbb{A}_+ \rightarrow \mathbb{R}$, with $PV(x, e)$ being denoted $PV(x|e)$.

Axiom DT2. For all $a, b \in \mathbb{R}$, all $x \in \mathbb{E}$, and all $i \in \mathbb{A}_+$,

$$PV(a \cdot x + b|i) = a \cdot PV(x|i) + b.$$

Axiom DT3. For all $i \in \mathbb{E}$ and $i' \in \mathbb{A}_+$ such that $i \cdot i' \in \mathbb{A}_+$, and all $x, x' \in \mathbb{A}$, if $PV(x|i \cdot i') = PV(x'|i \cdot i')$ then $PV(x \cdot i|i') = PV(x' \cdot i|i')$.

Axiom DT4. For all $i \in \mathbb{A}_+$, and all $x, x' \in \mathbb{A}$, we have $PV(x|i) \leq PV(x'|i)$.

Axiom DT5. For all $x, x', y \in \mathbb{A}$ and $i \in \mathbb{A}_+$, if $PV(x|i) = PV(x'|i)$ then $PV(x + y|i) = PV(x' + y|i)$.

We present our axioms now. Those labelled with primes are qualitatively weaker than their Dupré–Tipler analogs: Axioms 1' and 2'' assume nothing about taking complements, and Axiom 2' assumes less than Axiom 2. Axioms 1, 2 and 3 are numbered that way by analogy to DT1, DT2 and DT3, not because they are favoured over the weaker 1', 2'' or 2'.

Axiom 1'. Let \mathbb{E} be a family of fuzzy subsets of a set Ω , closed under intersections modelled by a t-norm \top . For all $\lambda \in [0, 1]$, the constant fuzzy set λ is in \mathbb{E} . Let \mathbb{E}_0 be a lower set of \mathbb{E} with respect

¹The same optimization may also be valid in Dupré and Tipler's original setting.

²The analogs of the axioms originally labelled 3 and 5 will not be needed, so we relabel them as 4 and 5 for clarity.

to inclusion, i.e. $A \subset B \in \mathbb{E}_0$ implies $A \in \mathbb{E}_0$. Set $\mathbb{E}_+ = \mathbb{E} \setminus \mathbb{E}_0$. There exists a function $p : \mathbb{E} \times \mathbb{E}_+ \rightarrow \mathbb{R}$, with $p(A, C)$ being denoted $p(A|C)$.

Axiom 1. Axiom 1' holds and \mathbb{E} is closed under complementation modelled by an involutive negation $c : [0, 1] \rightarrow [0, 1]$. The complement of $A \in \mathbb{E}$ will be denoted by A^c .

Interpretation: \mathbb{E} represents fuzzy events, $p(A|C)$ a measure of the event A conditional on the information given by C , \mathbb{E}_0 is a subfamily of negligible events under which $p(\cdot|C)$ is undefined. In the fuzzy setting it seems unjustified to assume *a priori* that the t-norm is the product.

Under Axiom 1, \mathbb{E} is a \top -clan in the sense of [13, Section 2]. Let \perp denote the t-conorm

$$a \perp b = c(c(a) \top c(b)), \quad a, b \in [0, 1] \quad (1)$$

which makes (\top, c, \perp) a *De Morgan triple*. Intersections, complements and unions of fuzzy sets will always be associated to a De Morgan triple.

Axiom 2''. For all $\lambda \in [0, 1]$, all $A \in \mathbb{E}$ and all $C \in \mathbb{E}_+$, we have

$$p(\lambda \cap A | C) = \lambda \top p(A | C)$$

and

$$p(1 | C) = 1.$$

Interpretation: Axiom 2'' establishes a commensurability between the p -scale and the membership scale, as it implies that the p -measure of the fuzzy event λ is the real number λ , and that p conforms to the handling of constant fuzzy sets 'preexistent to p '.

Axiom 2'. Axiom 2'' holds, moreover

$$p(A^c|C) = c(p(A|C)).$$

Axiom 2. Axiom 2' holds and $p(A|C) = 0$ whenever $C \in \mathbb{E}_+$ and $A \cap C \in \mathbb{E}_0$.³

Axiom 3. For all $A, A', B, C \in \mathbb{E}$ such that $B \cap C \in \mathbb{E}_+$, if $p(A|B \cap C) = p(A'|B \cap C)$ then $p(A \cap B|C) = p(A' \cap B|C)$.

Interpretation: That is a typical Bayesian view of how learning works.

Remark 2.1. A few comments. Note how Axiom 3 is weaker than the (often criticized) assumption in the Cox theorem that $p(A \cap B|C)$, as a function of three arguments, depends only on $p(A|B \cap C)$ and $p(B|C)$. In its turn, Dupré and Tipler's Axiom DT2 is already quite stronger than $p(A^c|C) = 1 - p(A|C)$, which is obtained only after some work in the Cox framework. In Axiom 2', we just take a general complement but will not aim at showing that c 'must' be $c(\lambda) = 1 - \lambda$, since doing that is not really natural for fuzzy sets. Finally, we surely acknowledge that in various uncertainty theories, like possibility and Dempster–Shafer, the mere assumption that $p(A^c|C)$ is a function of $p(A|C)$ is not natural.

³It has to be observed that the analog of this property does follow from the Dupré–Tipler axioms. There, $i \in \mathbb{A}_+$ if and only if $i \in \mathbb{A} \setminus \{0\}$. If $x \cdot i = 0 = 0 \cdot i$, from Axioms DT4 and DT2 one has $PV(x|i) = PV(0|i) = 0$.

3. The combination rule

Since we allow for t-norms other than the product, we speak of the 'combination rule' instead of the 'product rule'.

Theorem 3.1. Assume Axioms (1', 2'', 3). Then, for all $A, B, C \in \mathbb{E}$ such that $B \cap C \in \mathbb{E}_+$,

$$p(A \cap B | C) = p(A|B \cap C) \top p(B|C).$$

Proof. Consider the functions $f, g : \mathbb{E} \rightarrow [0, 1]$ given by $f(A) = p(A \cap B|C)$ and $g(A) = p(A|B \cap C)$. Notice that f is well-defined since $C \in \mathbb{E}_+$ (if $C \in \mathbb{E}_0$ then, by Axiom 1', also $B \cap C \in \mathbb{E}_0$, a contradiction). By Axiom 3,

$$g(A) = g(A') \Rightarrow f(A) = f(A').$$

Thus, there exists a function $f_{B,C}^* : [0, 1] \rightarrow [0, 1]$ such that $f(A) = f_{B,C}^*(g(A))$. Now consider the function $F : \mathbb{E} \times \mathbb{E} \times \mathbb{E}_+ \rightarrow [0, 1]$ given by

$$F(A, B, C) = p(A \cap B|C).$$

For any fixed $B \in \mathbb{E}$ and $C \in \mathbb{E}_+$ such that $B \cap C \in \mathbb{E}_+$, the reasoning above gives $F(A, B, C) = f_{B,C}^*(g(A))$. Therefore, there exists a function $F^* : [0, 1] \times \mathbb{E} \times \mathbb{E}_+ \rightarrow [0, 1]$ such that

$$F(A, B, C) = F^*(p(A|B \cap C), B, C) \quad (2)$$

whenever $B \cap C \in \mathbb{E}_+$. Indeed, it suffices to take

$$F^*(\lambda, B, C) = \begin{cases} f_{B,C}^*(\lambda), & B \cap C \in \mathbb{E}_+ \\ 0, & B \cap C \in \mathbb{E}_0. \end{cases}$$

From Axiom 2'', for any $D \in \mathbb{E}_+$ and $\lambda \in [0, 1]$,

$$p(\lambda|D) = \lambda \top p(1|D) = \lambda \top 1 = \lambda.$$

Applying (2) with $A = \lambda$,

$$\begin{aligned} F^*(\lambda, B, C) &= F^*(p(\lambda|B \cap C), B, C) \\ &= F(\lambda, B, C) = p(\lambda \cap B|C) = \lambda \top p(B|C), \end{aligned}$$

where the last identity uses Axiom 2''. Then, for any $A \in \mathbb{E}$,

$$\begin{aligned} p(A|B \cap C) &= F(A, B, C) = F^*(p(A|B \cap C), B, C) \\ &= p(A|B \cap C) \top p(B|C). \end{aligned}$$

□

When complementation enters the picture, one obtains the following dual result.

Theorem 3.2. Assume Axioms (1, 2', 3). For all $A, B, C \in \mathbb{E}$ such that $A^c \cap C \in \mathbb{E}_+$,

$$p(A \cup B|C) = p(A|C) \perp p(B|A^c \cap C).$$

Proof. Since (\top, c, \perp) is a De Morgan triple, with Axiom 2' and Theorem 3.1 we have

$$\begin{aligned} p(A \cup B|C) &= c(p((A \cup B)^c|C)) = c(p(A^c \cap B^c|C)) \\ &= c(p(A^c|C) \top p(B^c|A^c \cap C)) \\ &= c(p(A^c|C) \perp c(p(B^c|A^c \cap C))) \\ &= p(A|C) \perp p(B|A^c \cap C). \end{aligned}$$

□

Remark 3.1. Taking $c(\lambda) = 1 - \lambda$, $\top = \cdot$ and crisp events, one obtains the formula

$$P(A \cup B|C)$$

$$= P(A|C) + P(B \cap A^c|C) - P(A|C)P(B \cap A^c|C),$$

which is valid in traditional probability theory. This formula is quite clearer in the language of t-conorms.

4. Paradistributivity

Consider a triple (\top, c, \perp) , where \top and \perp are an arbitrary t-norm and an arbitrary t-conorm. Let us introduce a notion which will be central for the rest of the paper.

We will say that \top is *paradistributive with respect to c over \perp at $\lambda \in [0, 1]$* if

$$(\lambda \top \mu) \perp (\lambda \top c(\mu)) = \lambda$$

for all $\mu \in [0, 1]$. We will say *paradistributive* if it is paradistributive at all $\lambda \in [0, 1]$. For brevity, we will also say that the triple is *paradistributive* or *paradistributive at λ* .

The additivity results obtained in the next section when the triple (\top, c, \perp_L) is paradistributive should be contrasted to the behaviour induced by the paradistributivity of (\top, c, \perp) for \perp the De Morgan companion of \top and c .

Theorem 4.1. *Assume Axioms (1, 2', 3). For all $A, B \in \mathbb{E}$ and $C \in \mathbb{E}_+$, if (\top, c, \perp) is paradistributive at $p(A^c|C)$ then*

$$p(A|C) = p(A \cup B|C) \top p(A \cup B^c|C).$$

Dually, if (\top, c, \perp) is paradistributive at $p(A|C)$ then

$$p(A|C) = p(A \cap B|C) \perp p(A \cap B^c|C).$$

Proof. As seen in the proof of Theorem 3.2,

$$p(A \cup B|C) = c(p(A^c|C) \top p(B^c|A^c \cap C)).$$

Using Axioms 1 and 2',

$$c(p(A \cup B|C)) = p(A^c|C) \top c(p(B|A^c \cap C)),$$

whence

$$c(p(A \cup B|C) \perp [p(A^c|C) \top p(B|A^c \cap C)])$$

$$\begin{aligned} &= [p(A^c|C) \top c(p(B|A^c \cap C))] \perp [p(A^c|C) \top p(B|A^c \cap C)] \\ &= p(A^c|C) \end{aligned}$$

by the paradistributivity at $p(A^c|C)$.

Therefore, with Axiom 2' and Theorem 3.1,

$$\begin{aligned} p(A|C) &= c(c(p(A \cup B|C) \perp [p(A^c|C) \top p(B|A^c \cap C)])) \\ &= p(A \cup B|C) \top c(p(A^c|C) \top p(B|A^c \cap C)) \\ &= p(A \cup B|C) \top c(p(A^c \cap B|C)) \\ &= p(A \cup B|C) \top p(A \cup B^c|C). \end{aligned}$$

To obtain the second formula, we reason as follows (paradistributivity at $p(A|C)$ is used in the second step).

$$\begin{aligned} p(A|C) &= c(p(A^c|C)) = c(p(A^c \cup B|C) \top p(A^c \cup B^c|C)) \\ &= c(c(p(A \cap B^c|C)) \top c(p(A \cap B|C))) \\ &= p(A \cap B^c|C) \perp p(A \cap B|C). \end{aligned}$$

□

Remark 4.1. One may define paradistributivity of \perp over \top analogously. It can be checked that, if (\top, c, \perp) is a De Morgan triple, then \perp paradistributes over \top at μ if and only if \top paradistributes over \perp at $c(\mu)$. Thus Theorem 4.1 holds for every $A \in \mathbb{E}$ such that \perp is paradistributive over \top at $p(A|C)$ (first statement) or $p(A^c|C)$ (second statement).

At this point, we ignore whether paradistributive (at every point) De Morgan triples exist. An application of Theorem 4.1 is as follows. The *equilibrium* of c is denoted by e (i.e. $c(e) = e$). Let \wedge and \vee denote the minimum t-norm and the maximum t-conorm, respectively.

Corollary 4.2. *Assume Axioms (1, 2', 3). Assume further $\top = \wedge$. Let e be the equilibrium of c . For all $A, B \in \mathbb{E}$ and $C \in \mathbb{E}_+$,*

$$p(A|C) = \begin{cases} p(A \cup B|C) \wedge p(A \cup B^c|C), & p(A|C) \geq e \\ p(A \cap B|C) \vee p(A \cap B^c|C), & p(A|C) \leq e. \end{cases}$$

Proof. Let us check that the De Morgan triple (\wedge, c, \vee) is paradistributive at every $\lambda \in [0, e]$. Indeed, using the distributivity of \wedge over \vee , for any $\mu \in [0, 1]$

$$(\lambda \wedge \mu) \vee (\lambda \wedge c(\mu)) = \lambda \wedge (\mu \vee c(\mu)).$$

To prove that the right-hand side equals λ , it is enough to show that $\lambda \leq \mu \vee c(\mu)$. If $\mu \geq e$, then

$$\mu \vee c(\mu) \geq \mu \geq e \geq \lambda.$$

And if $\mu \leq e$, then $c(\mu) \geq c(e) = e$ and the same reasoning applies.

Now the result follows from Theorem 4.1 by noting that $p(A|C) \geq e$ if and only if $p(A^c|C) \leq c(e) = e$, i.e. if (\wedge, c, \vee) is paradistributive at $p(A^c|C)$. □

Remark 4.2. As Corollary 4.2 shows, the formula $p(A^c|C) = c(p(A|C))$ in Axiom 2' has a probabilistic flavor but is not enough to enforce additivity. This case is reminiscent of a necessity measure for p -big events and of a possibility measure for p -small events. This behaviour brings uninorms to mind.

5. Additivity

In this section, we show that Axioms 1-3 are enough to obtain conditional measures satisfying the classical properties of (finitely additive) probabilities.

The *standard complement* given by $c(\lambda) = 1 - \lambda$ will be denoted by s . Denote by \perp_L the *bounded sum* or *Lukasiewicz t-conorm* given by $\lambda \perp_L \mu = \min\{\lambda + \mu, 1\}$.

Theorem 5.1. *Assume Axioms (1, 2', 3). Assume further that $c = s$ and that (\top, s, \perp_L) is paradistributive. Then, for all $A, B, C \in \mathbb{E}$ such that $A^c \cap C \in \mathbb{E}_+$,*

$$\begin{aligned} p(A \cup B|C) &= p(A|C) \perp_L p(B \cap A^c|C) \\ &= p(A|C) + p(B \cap A^c|C). \end{aligned}$$

Proof. Notice that the union of fuzzy sets is still given by the De Morgan t-conorm \perp in (1), not \perp_L . Using Axiom 2', Theorem 3.1 and paradistributivity, we have

$$\begin{aligned} p(A \cup B|C) &= 1 - p(A^c \cap B^c|C) \\ &= 1 - [p(A^c|C) \top p(B^c|A^c \cap C)] \\ &= 1 - [p(A^c|C) \top (1 - p(B|A^c \cap C))] \\ &= 1 - [p(A^c|C) - p(A^c|C) \top p(B|A^c \cap C)] \\ &= 1 - p(A^c|C) + p(A^c \cap B|C) \\ &= p(A|C) + p(A^c \cap B|C). \end{aligned}$$

Since the sum in the right-hand side is $p(A \cup B|C)$, which is bounded above by 1,

$$p(A|C) \perp_L p(A^c \cap B|C) = p(A \cup B|C)$$

as well. \square

Theorem 5.2. *Assume Axioms (1, 2', 3). Assume further that (\top, c, \perp_L) is paradistributive. Then, for all $A, B, C \in \mathbb{E}$ such that $A \cap C \in \mathbb{E}_+$,*

$$p(A|C) = p(A \cap B|C) + p(A \cap B^c|C).$$

Proof. We have

$$\begin{aligned} &p(A \cap B|C) \perp_L p(A \cap B^c|C) \\ &= [p(A|C) \top p(B|A \cap C)] \perp_L [p(A|C) \top c(p(B|A \cap C))] \\ &= p(A|C). \end{aligned}$$

Since $p(A|C) \leq 1$ by Axiom 1', also

$$\begin{aligned} &p(A \cap B|C) \perp_L p(A \cap B^c|C) \\ &= p(A \cap B|C) + p(A \cap B^c|C). \end{aligned}$$

\square

Remark 5.1. The triple (\cdot, s, \perp_L) is paradistributive (as a consequence of the distributivity of \cdot over $+$). Is the product the only t-norm for which that holds?

Fuzzy sets $A, B \in \mathbb{E}$ will be called *disjoint* if $A \cap B = 0$ (remember that this is contingent on the choice of the t-norm \top).

Theorem 5.3. *Assume Axioms (1, 2', 3). Assume further that $c = s$ and that (\top, s, \perp_L) is paradistributive. If $A, B \in \mathbb{E}_+$ are disjoint and $A \cap C \in \mathbb{E}_+$, then*

$$p(A \cup B|C) = p(A|C) + p(B|C).$$

If $A_1, \dots, A_n \in \mathbb{E}$ are such that A_i is disjoint from $\bigcup_{j=i+1}^n A_j$ and $A_i \cap C \in \mathbb{E}_+$ for all $1 \leq i < n$, then

$$p\left(\bigcup_{i=1}^n A_i|C\right) = \sum_{i=1}^n p(A_i|C).$$

Proof. We claim that $B \cap A^c = B$ whenever A and B are disjoint. Indeed, assume that $B(\omega) \top A^c(\omega) = 0$ for all $\omega \in \Omega$. Then

$$[B(\omega) \top A(\omega)] \perp_L [B(\omega) \top s(A(\omega))] = B(\omega) \top s(A(\omega)).$$

By the paradistributivity, the left-hand side is $A(\omega)$, while the right-hand side is $(B \cap A^c)(\omega)$. That proves the claim.

Now, since $p(B \cap A^c|C) = p(B|C)$, it follows from Theorem 5.1 that

$$p(A \cup B|C) = p(A|C) + p(B|C)$$

whenever $A \cap C \in \mathbb{E}_+$.

The second part follows easily using induction. \square

Remark 5.2. Notice that Theorems 5.1 and 5.2 do not follow from Theorem 5.3 as they do in classical probability theory, since the events there (e.g. A and $B \cap A^c$) are not necessarily disjoint.

For the product t-norm, more can be said.

Theorem 5.4. *Assume Axioms (1, 2', 3). Assume further that $\top = \cdot$ and $c = s$. If $A_1, \dots, A_n \in \mathbb{E}$ are pairwise disjoint and $A_i \cap C \in \mathbb{E}_+$ for all $1 \leq i < n$, then*

$$p\left(\bigcup_{i=1}^n A_i|C\right) = \sum_{i=1}^n p(A_i|C).$$

Proof. Note that (\cdot, s, \perp_L) is paradistributive. Thus, in order to use Theorem 5.3 one just needs to check that pairwise disjointness implies the formally stronger disjointness property there.

Indeed, for any $\omega \in \Omega$,

$$\begin{aligned} (A_i \cap \bigcup_{j=i+1}^n A_j)(\omega) &= A_i(\omega) \cdot [1 - \prod_{j=i+1}^n (1 - A_j(\omega))] \\ &= A_i(\omega) - A_i(\omega) \prod_{j=i+1}^n (1 - A_j(\omega)) \end{aligned}$$

$$= A_i(\omega) - A_i(\omega) \cdot 1 = 0$$

because all the cross-products $A_i(\omega) \cdot A_j(\omega)$ are zero by the disjointness with respect to the product. \square

The reader must have observed the maybe uncomfortable assumptions like $A_i \cap C \in \mathbb{E}_+$. These are easily disposed of if one accepts Axiom 2.

Theorem 5.5. *Assume Axiom 2. Then, the conditions that $A^c \cap C \in \mathbb{E}_+$, $A \cap C \in \mathbb{E}_+$ and $A_i \cap C \in \mathbb{E}_+$ in Theorems 5.1, 5.2, 5.3 and 5.4 are unnecessary.*

Proof. We will prove it for Theorem 5.1, the others being similar and left to the reader.

We need to show

$$p(A \cup B|C) = p(A|C) + p(B \cap A^c|C) \quad (3)$$

when $C \in \mathbb{E}_+$ and $A^c \cap C \in \mathbb{E}_0$. By Axiom 2, $p(A^c|C) = 0$ and then

$$p(A|C) = c(p(A^c|C)) = c(0) = 1.$$

Moreover, since \mathbb{E}_0 is a lower set by Axiom 1', both $A^c \cap B \in \mathbb{E}_0$ and $A^c \cap B^c \in \mathbb{E}_0$ hold. By Axiom 2 again, $p(A^c \cap B|C) = 0$ and $p(A^c \cap B^c|C) = 0$. From the latter, Axiom 2' and the De Morgan property,

$$p(A \cup B|C) = c(p(A^c \cap B^c|C)) = 1.$$

Thence the left-hand side of (3) is 1, and the right-hand side is $1 + 0$. \square

Remark 5.3. Since Axioms 2 and even 2' are not valid in several uncertainty theories (e.g. the union of two sets with necessity 0 can have non-zero necessity), it seems worthwhile to explore the consequences of Axiom 2'' once it has been established that the theory contains a subtheory with the usual properties of probability. Accordingly, in Axiom 1' we have not assumed that \mathbb{E}_0 is closed under finite unions.

Remark 5.4. One can show that, for $\top = \cdot$ and under Axiom 2,

$$p(A|C) = \frac{p(A \cap C|1)}{p(C|1)}$$

whenever $p(C|1) > 0$.

6. Examples

6.1. Probability of fuzzy events

Let (Ω, \mathcal{A}, P) be a probability space. Let \mathbb{E} be the class of all fuzzy events (measurable fuzzy sets). Zadeh [19] proposed an extension of P from \mathcal{A} to \mathbb{E} , given by $P(A) = \int AdP$. Take

$$E_0 = \{C \in \mathbb{E} \mid P(C) = 0\}.$$

Define, for all $C \in \mathbb{E}_+$,

$$P(A|C) = \frac{P(A \cdot C)}{P(C)}$$

[19, p. 425]. Then the mapping $P(\cdot|C)$ satisfies Axioms 1, 2 and 3 for the choices $\top = \cdot$ and $c = s$.

6.2. Choquet capacities and conditional plausibility

One can extend the example above to an arbitrary Choquet capacity ν by replacing the Lebesgue integral against P by a Choquet integral against ν , i.e.

$$\nu(A) = \int_0^1 \nu(\{A \geq \alpha\}) d\alpha.$$

This has been used e.g. in [16]. Then the mapping $\nu(\cdot|C)$ given by

$$\nu(A|C) = \frac{\nu(A \cdot C)}{\nu(C)}$$

satisfies Axioms 1, 2'' and 3, and the property in Axiom 2 that $C \in \mathbb{E}_+$ and $A \cap C \in \mathbb{E}_0$ imply $\nu(A|C) = 0$, for the choices $\top = \cdot$ and $c = s$. But Axiom 2' fails, since the Choquet integral satisfies

$$\nu(A^c) = 1 - \bar{\nu}(A)$$

(where $\bar{\nu}$ is the dual capacity to ν) instead.

In the case when the capacity is a plausibility measure, Dempster's combination rule is obtained. The conclusion of Theorem 3.1, for the choice $\top = \cdot$, is singled out in [1] as one of the defining properties, property (ddd), of an axiomatized 'Dempster conditional plausibility'. Thus we can offer a justification for that property (and an extension to fuzzy events). Property (d) in that paper should probably be abandoned for fuzzy events, at it conflicts with non-idempotence. It is open whether, in that abstract setting, property (dd) is ensured by the Dupr e-Tipler construction, i.e. whether $p(\cdot|1)$ being a plausibility measure when restricted to crisp events implies that $p(\cdot|C)$ is so as well for all crisp $C \in \mathbb{E}_+$.

6.3. Conditional possibility

Let Π be a possibility measure defined on a τ -algebra (or 'ample field') \mathcal{A} . The relationship

$$\Pi(A \cap B) = \Pi(A|B) \top \Pi(B) \quad (4)$$

is often presented as a unifying definition of conditional possibility. It is interesting that this definition by analogy can be recovered from our axioms, which may provide a justification for it. To do so, set

$$\mathbb{E} = \{\lambda \cdot I_A \mid \lambda \in [0, 1], A \in \mathcal{A}\}$$

and extend Π to \mathbb{E} by

$$\Pi(\lambda \cdot I_A) = \lambda \top \Pi(A)$$

for some t-norm \top . Finally, take $\mathbb{E}_0 = \{0\}$.

Then, whenever $p : \mathbb{E} \times \mathbb{E}^+ \rightarrow [0, 1]$ satisfies Axioms 1', 2'' and 3, and agrees with Π in the sense that

$$p(A|1) = \Pi(A)$$

for all $A \in \mathbb{E}$, Theorem 3.1 with $C = 1$ implies

$$\Pi(A \cap B) = \Pi(B) \top p(A|B \cap 1) = \Pi(B) \top p(A|B).$$

In particular, for A, B crisp, one has (4).

7. Concluding remarks

A Bayesian would be likely to choose $\mathbb{E}_0 = \{0\}$ or even $\mathbb{E}_0 = \emptyset$. However, we see no reason to exclude *a priori* the possibility of a non-trivial family of negligible events upon which conditionalization is undefined (for instance, because one assumes they are not going to happen in practice).

The commensurability assumption $p(\lambda|1) = \lambda$ which follows from Axiom 2" is questionable. We do not make the claim that the axioms are 'true' or reject any conditional measure that fails one axiom or other. We just claim that studying them advances the understanding of the interplay between Bayesian arguments and their (usually taken for granted) logical assumptions.

Corollary 4.2 looks weird to the probabilistic eye, but it generalizes the following (Boolean) properties of crisp sets. If $C \subset A$, then $C \subset A \cup B$ and $C \subset A \cup B^c$. If $C \subset A \cap B$ or $C \subset A \cap B^c$, then $C \subset A$. This is obtained when p is 0-1 valued on crisp sets with

$$p(A|C) = \begin{cases} 1, & C \subset A \\ 0, & C \not\subset A, \end{cases}$$

as this p satisfies Axiom 3. Indeed, we have to show that, for a fixed C , if $B \cap C \subset A \iff B \cap C \subset A'$ then $C \subset A \cap B \iff C \subset A' \cap B$. The premise is equivalent to $A \cap C = A' \cap C$. By symmetry, to prove the conclusion it suffices to show $C \subset A \cap B \implies C \subset A' \cap B$. But if $C \subset A \cap B$ then $C = A \cap C = A' \cap C$, whence $C \subset A'$. That gives $C \subset A' \cap (A \cap B) \subset A' \cap B$.

Note that the construction in this paper *must* be founded on fuzzy events. It is not a more sophisticated variant of a construction available for Boolean events, as the key to the combination rule is to use events of the form $\lambda \cap A$. That is the main difference with the Cox approach.

The main divergence from Dupré-Tipler is that the structures demanded by Axioms DT1 and 1 are entirely different, whence non-additive solutions appear in our case. Moreover, note that their \mathbb{A}_+ is much smaller than \mathbb{E} : they only allow conditioning on idempotent elements which form a Boolean algebra, while elements of our \mathbb{E}_+ , which can be as big as $\mathbb{E} \setminus \{0\}$ or even \mathbb{E} , are typically non-idempotent and do not form a Boolean algebra. Finally, it should be checked whether Dupré and Tipler's plan can be carried out without invoking Axioms DT4 and DT5, since we have not needed analogs to them.

References

[1] G. Coletti, M. Mastroleo. Conditional belief functions: a comparison among different definitions. *Proceedings of the 7th Workshop on Uncertainty Processing*, paper no. 4, 2006. Available at <http://www.utia.cas.cz/files/mtr/articles/data/coletti.pdf>

[2] R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics* 14:1–13, 1946.

[3] M. J. Dupré, F. J. Tipler. New axioms for rigorous Bayesian probability. *Bayesian Analysis* 4:599–606, 2009.

[4] P. Goyal, K. H. Knuth. Quantum theory and probability theory: their relationship and origin in symmetry. *Symmetry* 3:171–206, 2011.

[5] F. Holik, M. Sáenz, A. Plastino. A discussion on the origin of quantum probabilities. *Annals of Physics* 340:293–310, 2014.

[6] D. Hunter. Uncertain reasoning using maximum entropy inference. *Proceedings of the 1st International Conference on Uncertainty in Artificial Intelligence*, pages 21–27, 1985.

[7] E. T. Jaynes. *Probability theory: the logic of science*. Cambridge Univ. Press, Cambridge, 2003.

[8] K. H. Knuth. Valuations on lattices: fuzzification and its implications. In: V. G. Kaburlasos, G. X. Ritter, editors, *Computational intelligence based on lattice theory*, pages 309–324. Springer, Berlin, 2007.

[9] K. H. Knuth. Measuring on lattices. In P. M. Goggans, C.-Y. Chan, editors, *Bayesian inference and maximum entropy methods in science and engineering*, pages 132–144. American Institute of Physics, Melville, 2009.

[10] B. Kosko. Fuzzy associative memory systems. In A. Kandel, editor, *Fuzzy expert systems*, pages 135–154. CRC, Boca Raton, 1992.

[11] J. Kühn, D. Mundici. De Finetti theorem and Borel states in $[0, 1]$ -valued algebraic logic. *International Journal of Approximate Reasoning* 46:605–616, 2007.

[12] D. Mundici. Bookmaking over infinite-valued events. *International Journal of Approximate Reasoning* 43:223–240, 2006.

[13] M. Navara. Tribes revisited. In U. Bodenhofer, B. de Baets, E. P. Klement, S. Saminger-Platz, editors, *LINZ 2009. The legacy of 30 seminars—where do we stand and where do we go*, pages 81–84. Johannes Kepler Universität, Linz, 2009.

[14] G. Pólya. Heuristic reasoning and the theory of probability. *American Mathematical Monthly* 48:450–465, 1941.

[15] J. Skilling. Probabilistic data analysis: an introductory guide. *Journal of Microscopy* 190:28–36, 1998.

[16] P. Terán. Centrality as a gradual notion: a new bridge between fuzzy sets and statistics. *International Journal of Approximate Reasoning* 52:1243–1256, 2011.

[17] K. S. Van Horn. Constructing a logic of plausible inference: a guide to Cox's theorem. *International Journal of Approximate Reasoning* 34:3–24, 2003.

[18] B. Weatherston. From classical to intuitionistic probability. *Notre Dame Journal of Formal Logic* 44:111–123, 2003.

- [19] L. A. Zadeh. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications* 23:421–427, 1968.