Belief functions induced by random fuzzy sets: A general framework for representing uncertain and fuzzy evidence

Thierry Denœux[a,b,c]

[a]*Université de technologie de Compiègne, CNRS*
UMR 7253 Heudiasyc, Compiègne, France
[b]*Shanghai University, UTSEUS, Shanghai, China*
[c]*Institut universitaire de France, Paris, France*

Abstract

We revisit Zadeh's notion of "evidence of the second kind" and show that it provides the foundation for a general theory of epistemic random fuzzy sets, which generalizes both the Dempster-Shafer theory of belief functions and possibility theory. In this perspective, Dempster-Shafer theory deals with belief functions generated by random sets, while possibility theory deals with belief functions induced by fuzzy sets. The more general theory allows us to represent and combine evidence that is both uncertain and fuzzy. We demonstrate the application of this formalism to statistical inference, and show that it makes it possible to reconcile the possibilistic interpretation of likelihood with Bayesian inference.

Keywords: Dempster-Shafer theory, evidence theory, possibility theory, fuzzy mass functions, uncertain reasoning, likelihood, estimation, prediction.

1. Introduction

The theory of belief functions developed by Dempster [7] and Shafer [42] is a theory of uncertain reasoning that puts the emphasis on the concept of *evidence*. It is based on the representation of elementary pieces of evidence by belief functions (defined as completely monotone set functions) and on their combination by an operator called the product-intersection rule, or Dempster's rule of combination. A belief function can be constructed by comparing a piece evidence to a scale of canonical examples such as randomly coded messages, whose meanings are determined by chance [43]. A belief function on a set Θ can be seen as being induced by a multi-valued mapping from a probability space to Ω; it is mathematically equivalent to a random set [7, 37]. As rational beliefs are essentially determined by evidence, the Dempster-Shafer (DS) theory can be regarded as a general framework for reasoning with uncertainty [13].

Shortly after the introduction of DS theory, Zadeh independently proposed another formalism, called *Possibility Theory* [57], in which the concept of "fuzzy restriction" plays a

central role (see, e.g., [14] for a recent review). A fuzzy restriction is typically imposed on a variable X taking values in a set Θ by a statement of the form "X is $\widetilde{F}$", where $\widetilde{F}$ is a fuzzy subset of Θ. For instance, the statement "John is young" acts as a flexible constraint on the age of John. Zadeh proposed to identify the membership function of the fuzzy set $\widetilde{F}$ with a *possibility distribution*: for any $\theta \in \Theta$, $\widetilde{F}(\theta)$ is then both the degree of membership of θ to the fuzzy set $\widetilde{F}$, and the degree of possibility that X takes value θ, knowing that "X is $\widetilde{F}$". The possibility of a subset $A \subseteq \Theta$ is then the supremum of the degrees of possibility $\pi(\theta)$ for all $\theta \in \Theta$.

The mathematical connections between the two theories were soon pointed out [20], namely: a possibility measure corresponds to a belief function induced by a multi-valued mapping that assigns probabilities only to a collection of nested subsets. Such a belief function is said to be *consonant*. From this point of view, the formalism of belief functions is more general than that of possibility theory. However, Dempster's rule does not fit in possibility theory as it does not preserve consonance: the combination of two possibility measures by Dempster's rule is not a possibility measure. Possibility theory has its own conjunctive and disjunctive combination operators based on triangular norms and conorms [23]: consequently, it is not a special case of DS theory, but a stand-alone framework. An open question is whether DS and possibility theories should be regarded are two "competing" models of uncertainty, which can be used interchangeably to reason with partial information, or whether they should rather be considered as complementary, each theory being suited to represent different kinds of uncertain evidence. If one adopts the latter view, as I do in this paper, it makes sense to search for a more general theory that encompasses belief functions, possibility measures, and their combination operators as special cases. Interestingly, elements of such a theory already exist, but they have received relatively little attention until now.

The idea of combining DS and possibility theories can first be found in an early paper by Zadeh [58]. Zadeh calls "evidence of the second kind" a pair $(X, \Pi_{(Y|X)})$ in which X is a discrete random variable on a set Ω and $\Pi_{(Y|X)}$ a collection of "conditioned π-granules", i.e., conditional possibility distributions of Y given $X = x$, for all $x \in \Omega$. The probability distribution on X and the conditional possibility distributions together define a *mixture of possibility measures*. If the random variable X is certain, then the mixture has only one component and it boils down to a single possibility measure representing fuzzy evidence. If the conditional possibility distributions take values in $\{0, 1\}$, we essentially get a DS belief function, representing uncertain evidence. In general, Zadeh's concept of evidence of the second kind makes it possible to represent evidence that is both fuzzy and uncertain, and it lays down the foundation of a general theory of uncertainty encompassing possibility theory and DS theory as special cases. This theory, only sketched by Zadeh in [58] and mostly overlooked in later work, is further explored in this paper.

The mathematical structure $(X, \Pi_{(Y|X)})$ is actually related to the notion of *random fuzzy set* or *fuzzy random variable*, a concept that also appeared at the end of the 1970's [26, 33, 34]. Random fuzzy sets have been used with different interpretations that differ from Zadeh's evidence of the second kind, such as a model of a random mechanism for generating

2

fuzzy data [41, 27], or a representation of imprecise information about the true probability distribution associated with a random experiment [32, 4]. Baudrit et al [3] use random fuzzy sets to propagate probabilistic and possibilistic uncertainty in risk analysis from an imprecise probability perspective. Couso and Sánchez [5] propose a framework based on a known probability measure on a set Ω_1 and a family of conditional possibility measures $\{\Pi(\cdot|\omega) : \omega \in \Omega_1\}$ on a set Ω_2, similar to Zadeh's "conditioned π-granules". Independently from the literature on fuzzy random variables, a few contributions on "fuzzy belief structures", defined as DS mass functions with fuzzy focal sets, were also published in the 1980's and early 1990's [29, 51, 53]. Applications to classification, regression and missing data imputation were presented, respectively, in [17], [39, 40] and [38]. However, in these papers, the formalism of fuzzy belief structures was seen merely as a fuzzy generalization of DS mass functions, and not as a general concept encompassing DS and possibility theories as special cases, a perspective that is adopted in this paper.

A very important problem for which a general theory of epistemic random fuzzy sets can be useful is *statistical inference*. In [42], Shafer proposed to treat the relative likelihood function as the contour function or a consonant belief function. This view is quite appealing and it was further justified axiomatically in [12]. However, it was later rejected by Shafer [44] because it is not consistent with Dempster's rule of combination: the belief function induced by a joint random sample composed of two independent samples is not the orthogonal sum of the belief functions induced by each of the two samples considered separately. Rather, relative likelihood functions induced by independent samples must be multiplied and renormalized, an operation that is consistent with a possibilistic interpretation of the likelihood [46, 1]. Yet, the limited expressiveness of possibility theory alone does not allow it to represent probability distributions and their combination with likelihoods to yield posterior distributions. The generalized setting proposed in this paper allows us to resolve these difficulties, by viewing the likelihood function as defining a mass function with a single fuzzy focal set. The resulting model still generalizes Bayesian inference as did Shafer's original model, while being consistent with Dempster's rule of combination. To highlight the basic principles without being distracted by mathematical intricacies, we will assume the variables to be defined on finite domains throughout this paper. The case of infinite domains will be addressed in a companion paper.

The rest of the paper is organized as follows. DS and possibility theories are first recalled in Section 2, where some new definitions and results are also given. The more general random fuzzy set model is then exposed in Section 3. Finally, the application to statistical inference is addressed in Section 4, and Section 5 concludes the paper.

2. DS and possibility theories

In this section, we consider a variable $\boldsymbol{\theta}$ taking values in a finite set Θ, and we review different classical models for representing and combining evidence about $\boldsymbol{\theta}$. The case of logical evidence is first recalled in Section 2.1. This simple case can be generalized to *uncertain* evidence, leading to DS theory, or to *fuzzy* evidence, resulting in possibility theory. These two theories are reviewed, respectively, in Sections 2.2 and 2.3.

2.1. Logical evidence

Let F be a subset of Θ, and assume that we receive a piece of evidence that us that $\boldsymbol{\theta} \in F$ for sure, and nothing more. We call such evidence "logical" because it consists of a proposition that is known to be true. Given this evidence, a proposition "$\boldsymbol{\theta} \in A$" for some $A \subseteq \Theta$ is *possible* if and only if $F \cap A \neq \emptyset$, and it is *certain* if and only if $F \subseteq A$. We can define two set functions Π_F and N_F from the power set 2^Θ to $\{0, 1\}$, as

$$\Pi_F(A) := I(F \cap A \neq \emptyset)$$

and

$$N_F(F) := I(F \subseteq A),$$

where $I(\cdot)$ is the indicator function, which returns 1 if its argument is true, and 0 otherwise [24]. We can remark that $\Pi_F(A) = 1$ if and only if there is some $\theta \in A$ that belongs to F. We can thus write, equivalently,

$$\Pi_F(A) = \max_{\theta \in A} F(\theta), \tag{1}$$

where $F(\cdot)$ denotes the characteristic function of set F. The possibility value of $\{\theta\}$, denoted by $\pi_F(\theta)$, is

$$\pi_F(\theta) := \Pi_F(\{\theta\}) = F(\theta). \tag{2}$$

Furthermore, the proposition $F \subseteq A$ is equivalent to $F \cap A^c = \emptyset$, where A^c denotes the complement of A. Consequently, we have

$$N_F(A) = 1 - \Pi_F(A^c). \tag{3}$$

It is clear that, for any two subsets A and B of Θ, the following equalities hold:

$$\Pi_F(A \cup B) = \Pi_F(A) \vee \Pi_F(B) \tag{4}$$

and

$$N_F(A \cap B) = N_F(A) \wedge N_F(B), \tag{5}$$

where $\vee$ and $\wedge$ denote, respectively, the maximum and minimum operators. Functions Π_F and N_F can be called, respectively, *Boolean possibility* and *necessity* measures, and function $\pi_F : \Theta \to \{0, 1\}$ can be called a *Boolean possibility distribution*.

A subset $A \subseteq \Theta$ is possible if at least one element $\theta \in A$ is possible (i.e., belongs to F). This is captured by function Π_F. We can also define the stronger notion of *guaranteed possibility* [24]: $A \subseteq \Theta$ is "guaranteed possible" if *all* its elements are possible, i.e., if $A \subseteq F$. This notion is captured by the guaranteed possibility function

$$\Delta_F(A) := I(A \subseteq F) = \min_{\theta \in A} F(\theta). \tag{6}$$

We can also define a dual "potential certainty" function, $\nabla_F(A) = 1 - \Delta_F(A^c)$, which equals one if and only if there is some θ outside A that is not possible, i.e., if $A \cup F \neq \Theta$.

If we now have two pieces of evidence telling us that $\boldsymbol{\theta} \in F$ for sure and $\boldsymbol{\theta} \in G$ for sure, and if we consider that both sources can be trusted, then we can infer that $\boldsymbol{\theta} \in F \cap G$. The combined Boolean possibility distribution is, thus,

$$\pi_{F \cap G}(\theta) := \pi_F(\theta) \wedge \pi_G(\theta). \tag{7}$$

If, on the other hand, we only consider that at least one of the two sources can be trusted, then we can infer that $\boldsymbol{\theta} \in F \cup G$, and the combined Boolean possibility distribution is

$$\pi_{F \cup G}(\theta) := \pi_F(\theta) \vee \pi_G(\theta). \tag{8}$$

This simple model can be extended in two ways, by allowing the piece of information to be uncertain, or fuzzy. These two extensions correspond, respectively, to Dempster-Shafer (DS) and Possibility theories, which are recalled in Sections 2.2 and 2.3 below.

2.2. Uncertain evidence: DS theory

Let us now assume that we receive a piece of evidence that can be interpreted in different ways, with given probabilities [42, 13]. Let Ω be the set of interpretations, assumed to be finite. If interpretation $\omega \in \Omega$ holds, then the evidence tells us that $\boldsymbol{\theta}$ belongs to some nonempty subset $\Gamma(\omega) \subseteq \Theta$, and nothing more. We further assume that we can assess probabilities on Ω, on which we define a probability measure P. The tuple $(\Omega, 2^{\Omega}, P, \Gamma)$, where Γ is a mapping from Ω to 2^{Θ} is a *random set*. We define the corresponding *mass function* $m : 2^{\Theta}$ to $[0, 1]$ as

$$m(A) := P\left(\{\omega \in \Omega : \Gamma(\omega) = A\}\right),$$

for all nonempty subset $A \subseteq \Theta$, and $m(\emptyset) = 0$. A subset F such that $m(F) > 0$ is called a *focal set* of m. We denote the focal sets by $F_1, \ldots, F_f$, and their masses by $m_i := m(F_i)$ for $i = 1, \ldots, f$. A mass function is said to be *logical* if it has only one focal set, and *Bayesian* if all of its focal sets are singletons.

If interpretation ω holds, we know that $\boldsymbol{\theta} \in \Gamma(\omega)$. The possibility and necessity that $\boldsymbol{\theta} \in A$ are then, respectively, $\Pi_{\Gamma(\omega)}(A)$ and $N_{\Gamma(\omega)}(A)$. We can then compute the *expected possibility* and the *expected necessity* of the proposition "$\boldsymbol{\theta} \in A$" as, respectively,

$$Pl_m(A) = \sum_{\omega \in \Omega} P(\{\omega\}) \Pi_{\Gamma(\omega)}(A) = \sum_{i=1}^{f} m_i \Pi_{F_i}(A) = \sum_{i=1}^{f} m_i I(F_i \cap A \neq \emptyset) \tag{9}$$

and

$$Bel_m(A) = \sum_{\omega \in \Omega} P(\{\omega\}) N_{\Gamma(\omega)}(A) = \sum_{i=1}^{f} m_i N_{F_i}(A) = \sum_{i=1}^{f} m_i I(F_i \subseteq A). \tag{10}$$

These two functions are called, respectively, *plausibility* and a *belief* functions. If m has only one focal set F, it is clear that functions Pl_m and Bel_m boil down, respectively, to the Boolean possibility and necessity functions Π_F and N_F reviewed in Section 2.1. In

5

the general case, they are mixtures of such functions. The restriction of function Pl_m to singletons, denoted as $pl_m(\theta) := Pl_m(\{\theta\})$, is called the *contour function* associated to m. It is equal to function π_F when m has only one focal set F.

Functions Pl_m and Bel_m are linked by the duality relation $Bel_m(A) = 1 - Pl_m(A^c)$, which is a direct consequence of (3). Shafer [42] shows that a mapping $Bel : 2^\Theta \to [0,1]$ can be written in the form (10) if and only if $Bel(\emptyset) = 0$, $Bel(\Theta) = 1$, and Bel is completely monotone, i.e.,

$$Bel\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i\in I} A_i\right) \tag{11}$$

for any $k \geq 2$ and any collection $A_1, \dots, A_k$ of subsets of Θ.

In addition to the expected possibility and necessity, we can also compute the *expected guaranteed possibility* of the proposition "$\boldsymbol{\theta} \in A$" as

$$Q_m(A) = \sum_{\omega\in\Omega} P(\{\omega\}) \Delta_{\Gamma(\omega)}(A) = \sum_{i=1}^{f} m_i \Delta_{F_i}(A) = \sum_{i=1}^{f} m_i I(A \subseteq F_i). \tag{12}$$

Function Q_m is called the *commonality function* induced by m [42]; it plays an important role in DS theory, as will be shown below. The dual function $\nabla_m(A) = 1 - Q_m(A^c)$ could also be defined, but its interpretation in DS theory is less clear. Functions m, Bel_m, Pl_m and Q_m are in one-to-one correspondence and any one of them allows us to recover the other three. They can thus be considered as different facets of the same information.

Dempster's rule. Let us now assume that we have two mass functions m_1 and m_2 on Θ induced by *independent* pieces of evidence, i.e., by independent random sets $(\Omega_1, 2^{\Omega_1}, P_1, \Gamma_1)$ and $(\Omega_2, 2^{\Omega_2}, P_2, \Gamma_2)$. We further assume that both pieces of evidence are reliable, i.e., if the pair of interpretations $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ holds, then we know for sure that $\boldsymbol{\theta} \in \Gamma_\cap(\omega_1, \omega_2) := \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$, provided that $\Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) \neq \emptyset$. To compute the probability that a particular event in $\Omega_1 \times \Omega_2$ holds, we need to compute the product measure $P := P_1 \otimes P_2$ and to condition it on the event

$$\Theta^* := \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) \neq \emptyset\}. \tag{13}$$

The *orthogonal sum* [42] of m_1 and m_2 is then defined, for all $A \in 2^\Theta$, as

$$(m_1 \oplus m_2)(A) := P(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = A\} \mid \Theta^*) \tag{14a}$$

$$= \frac{P(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = A\} \cap \Theta^*)}{P(\Theta^*)} \tag{14b}$$

$$= \begin{cases} 0 & \text{if } A = \emptyset \\ \dfrac{\sum_{F\cap G=A} m_1(F) m_2(G)}{\sum_{F\cap G\neq\emptyset} m_1(F) m_2(G)} & \text{otherwise,} \end{cases} \tag{14c}$$

which is well defined on the condition that $P(\Theta^*) > 0$. The binary operation $\oplus$ is called *Dempster's rule*. It is commutative and associative. The quantity

$$\kappa := 1 - P(\Theta^*) = \sum_{F \cap G = \emptyset} m_1(F) m_2(G)$$

is called the *degree of conflict* between m_1 and m_2.

The following two propositions are important in practice [42].

Proposition 1. *Let m_1 and m_2 be two mass functions on Θ. The commonality function $Q_{m_1 \oplus m_2}$ corresponding to the orthogonal sum $m_1 \oplus m_2$ is proportional to the product of the commonality functions Q_{m_1} and Q_{m_2} corresponding to m_1 and m_2:*

$$Q_{m_1 \oplus m_2} = \frac{1}{1 - \kappa} Q_{m_1} Q_{m_2}.$$

As $Q_m(\{\theta\}) = pl_m(\theta)$ for all $\theta \in \Theta$, we also have $pl_{m_1 \oplus m_2} = (1 - \kappa)^{-1} pl_{m_1} pl_{m_2}$.

Proposition 2. *Let m_1 be a mass function and let m_2 be a Bayesian mass function. Then, the mass function $m_1 \oplus m_2$ is Bayesian; it is defined as*

$$(m_1 \oplus m_2)(\{\theta\}) = \frac{pl_{m_1}(\theta) m_2(\{\theta\})}{\sum_{\theta' \in \Theta} pl_{m_1}(\theta') m_2(\{\theta'\})}$$

for all $\theta \in \Theta$.

As mentioned above, Dempster's rule is based on the assumption that both pieces of evidence are reliable. We can also weaken this assumption and only assume that *at least one of them* is reliable [22, 47]. If the pair of interpretations $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ holds, we can then deduce that $\boldsymbol{\theta} \in \Gamma_\cup(\omega_1, \omega_2) := \Gamma_1(\omega_1) \cup \Gamma_2(\omega_2)$. Still assuming the two pieces of evidence to be independent, we get the combined mass function $m_1 \uplus m_2$ defined, for all $A \in 2^\Theta$, as

$$(m_1 \uplus m_2)(A) := P(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cup(\omega_1, \omega_2) = A\}) \tag{15a}$$

$$= \sum_{F \cup G = A} m_1(F) m_2(G). \tag{15b}$$

We note that normalization is not needed in this case, as $\Gamma_\cup(\omega_1, \omega_2)$ cannot be empty as long as $\Gamma_1(\omega_1)$ and $\Gamma_2(\omega_2)$ are nonempty.

Degree of belief in a fuzzy event. The notions of a fuzzy event and its probability were first defined by Zadeh [55]. Given a probability space $(\Theta, \mathcal{A}, P)$, a *fuzzy event* is a fuzzy subset $\widetilde{A}$ of Θ with measurable membership function, and the probability of $\widetilde{A}$ is defined as the expectation of its membership function. When Θ is finite, as assumed throughout his paper, and $\mathcal{A} = 2^\Theta$, the measurability of $\widetilde{A}$ is always satisfied and we have

$$P(\widetilde{A}) := \sum_{\theta \in \Theta} P(\{\theta\}) \widetilde{A}(\theta), \tag{16}$$

where $\widetilde{A}(\theta)$ denotes the degree of membership of θ in the fuzzy set $\widetilde{A}$. It can be shown [19] that $P(\widetilde{A})$ can also be written as:

$$P(\widetilde{A}) = \int_0^1 P(^\alpha\widetilde{A})d\alpha, \tag{17}$$

where $^\alpha\widetilde{A} = \{\theta \in \Theta : \widetilde{A}(\theta) \geq \alpha\}$ is the α-cut of $\widetilde{A}$.

Smets [45] extended this definition to the case where uncertainty on Θ is defined by a mass function m. He defined the *degrees of belief and plausibility* of fuzzy event $\widetilde{A}$ as, respectively, the lower and upper expectations of its membership function:

$$Bel_m(\widetilde{A}) = \sum_{i=1}^f m_i \min_{\theta \in F_i} \widetilde{A}(\theta) \tag{18a}$$

$$Pl_m(\widetilde{A}) = \sum_{i=1}^f m_i \max_{\theta \in F_i} \widetilde{A}(\theta). \tag{18b}$$

Similarly to (17), we have

$$Bel_m(\widetilde{A}) = \int_0^1 Bel_m(^\alpha\widetilde{A})d\alpha \tag{19a}$$

$$Pl_m(\widetilde{A}) = \int_0^1 Pl_m(^\alpha\widetilde{A})d\alpha. \tag{19b}$$

We note that $Bel_m(\widetilde{A})$ and $Pl_m(\widetilde{A})$ are the Choquet integrals of $\widetilde{A}(\cdot)$ with respect to Bel_m and Pl_m, respectively. It is clear that definition (18) coincides with (16) when m is Bayesian. Let $\mathcal{F}(\Theta)$ be the set of all fuzzy subsets of Θ. It becomes a lattice when equipped with fuzzy intersection and union defined, respectively, as

$$(\widetilde{A} \wedge \widetilde{B})(\theta) = \widetilde{A}(\theta) \wedge \widetilde{B}(\theta) \tag{20a}$$

and

$$(\widetilde{A} \vee \widetilde{B})(\theta) = \widetilde{A}(\theta) \vee \widetilde{B}(\theta) \tag{20b}$$

for all $\theta \in \Theta$, where $\vee$ and $\wedge$ denote, respectively, the minimum and the maximum. As shown by Smets [45], the mapping Bel from $\mathcal{F}(\Theta)$ to $[0,1]$ defined by (18) is a belief function on $(\mathcal{F}(\Theta), \wedge, \vee)$, i.e., it verifies the following inequalities for any $k \geq 2$ and countable collection $\widetilde{A}_1, \ldots, \widetilde{A}_k$ of fuzzy subsets of Θ:

$$Bel_m\left(\bigvee_{i=1}^k \widetilde{A}_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Bel_m\left(\bigwedge_{i \in I} \widetilde{A}_i\right). \tag{21}$$

Although Smets did not consider extending the commonality function to fuzzy events, this can also be done as follows:

$$Q_m(\widetilde{A}) := \int_0^1 Q(^\alpha \widetilde{A}) d\alpha \tag{22a}$$

$$= \int_0^1 \left(\sum_{i=1}^f m_i I(^\alpha \widetilde{A} \subseteq F_i) \right) d\alpha \tag{22b}$$

$$= \sum_{i=1}^f m_i \int_0^1 I(^\alpha \widetilde{A} \subseteq F_i) d\alpha \tag{22c}$$

$$= \sum_{i=1}^f m_i \left(1 - \max_{\theta \notin F_i} \widetilde{A}(\theta) \right). \tag{22d}$$

2.3. Fuzzy evidence: possibility theory

As recalled in Section 1, possibility theory introduced by Zadeh [57] extends the simple model outlined in Section 2.1 by considering statements of the form "$\boldsymbol{\theta}$ is $\widetilde{F}$", where $\widetilde{F}$ si a normal fuzzy subset of Θ, i.e., a fuzzy subset verifying $\widetilde{F}(\theta) = 1$ for some $\theta \in \Theta$ (see also [23, 24, 14]). Such fuzzy statements are good representations of evidence that can be fully trusted but that does not have a precise meaning because, e.g., it is conveyed through natural language. For instance, I may know from a fully reliable witness that "John is old". Zadeh [57] defines the *possibility measure* $\Pi_{\widetilde{F}} : 2^\Theta \to [0,1]$ induced by the piece of information "$\boldsymbol{\theta}$ is $\widetilde{F}$" as

$$\Pi_{\widetilde{F}}(A) := \max_{\theta \in A} \widetilde{F}(\theta), \tag{23}$$

which generalizes (2). The dual *necessity measure* is defined as

$$N_{\widetilde{F}}(A) := 1 - \Pi_{\widetilde{F}}(A^c) = \min_{\theta \notin A} \left[1 - \widetilde{F}(\theta) \right] \tag{24}$$

for all $A \subseteq \Theta$, which extends (3). The corresponding *possibility distribution* is defined as

$$\pi_{\widetilde{F}}(\theta) := \Pi_{\widetilde{F}}(\{\theta\}) = \widetilde{F}(\theta)$$

for all $\theta \in \Theta$. It is, thus, numerically equal to $\widetilde{F}$.

It can easy be seen that

$$\Pi_{\widetilde{F}}(A) = \int_0^1 \Pi_{\alpha \widetilde{F}}(A) d\alpha \tag{25a}$$

and

$$N_{\widetilde{F}}(A) = \int_0^1 N_{\alpha \widetilde{F}}(A) d\alpha. \tag{25b}$$

Trivially, we have $\Pi_{\widetilde{F}}(\emptyset) = N_{\widetilde{F}}(\emptyset) = 0$, $\Pi_{\widetilde{F}}(\Theta) = N_{\widetilde{F}}(\Theta) = 1$, and $N_{\widetilde{F}}(A) \leq \Pi_{\widetilde{F}}(A)$ for all $A \subseteq \Theta$. Furthermore, the following equalities hold:

$$\Pi_{\widetilde{F}}(A \cup B) = \Pi_{\widetilde{F}}(A) \vee \Pi_{\widetilde{F}}(B) \tag{26a}$$

and

$$N_{\widetilde{F}}(A \cap B) = N_{\widetilde{F}}(A) \wedge N_{\widetilde{F}}(B) \tag{26b}$$

for all $A, B \subseteq \Theta$. These equalities generalize (4) and (5).

In addition to the possibility and necessity measures defined by (26), Dubois and Prade [24] also generalize the notion of *guaranteed possibility* (6) as

$$\Delta_{\widetilde{F}}(A) := \min_{\theta \in A} \widetilde{F}(\theta) = \int_0^1 \Delta_{\alpha \widetilde{F}}(A) d\alpha, \tag{27}$$

and that of *potential certainty* as

$$\nabla_{\widetilde{F}}(A) := 1 - \Delta_{\widetilde{F}}(A^c) = \max_{\theta \notin A} \left[1 - \widetilde{F}(\theta) \right] = \int_0^1 \nabla_{\alpha \widetilde{F}}(A) d\alpha. \tag{28}$$

The quantity $\Delta_{\widetilde{F}}(A)$ measures the extent to which *all* values in A are actually possible given the statement "$\boldsymbol{\theta}$ is $\widetilde{F}$", while $\nabla_{\widetilde{F}}(A)$ measures the extent to which *at least one* value θ outside A has a low degree of possibility. Clearly, the following equality holds for any subsets A and B of Θ:

$$\Delta_{\widetilde{F}}(A \cup B) = \Delta_{\widetilde{F}}(A) \wedge \Delta_{\widetilde{F}}(B). \tag{29}$$

Combination of possibility measures. Let us now assume that we receive two pieces of evidence that tell us that "$\boldsymbol{\theta}$ is $\widetilde{F}$" and "$\boldsymbol{\theta}$ is $\widetilde{G}$", where $\widetilde{F}$ and $\widetilde{G}$ are two normal fuzzy subsets of Θ. These fuzzy sets induce two possibility distributions $\pi_{\widetilde{F}}$ and $\pi_{\widetilde{G}}$. How should they be combined? If both sources are assumed to be reliable, then it makes sense to infer that "$\boldsymbol{\theta}$ is $\widetilde{F} \cap \widetilde{G}$", where $\widetilde{F} \cap \widetilde{G}$ denotes the intersection of fuzzy sets $\widetilde{F}$ and $\widetilde{G}$. There are, however, two difficulties. First, there are several possible definitions of fuzzy set intersection. Zadeh [54] defines the intersection of fuzzy sets $\widetilde{F}$ and $\widetilde{G}$ as

$$(\widetilde{F} \wedge \widetilde{G})(\theta) = \widetilde{F}(\theta) \wedge \widetilde{G}(\theta),$$

but he also proposes the following alternative definition:

$$(\widetilde{F} \cdot \widetilde{G})(\theta) = \widetilde{F}(\theta) \cdot \widetilde{G}(\theta).$$

It is clear that both definitions extend the usual set intersection. Later, these two definitions have been generalized as $(\widetilde{F} \cap_\top \widetilde{G})(\theta) := \widetilde{F}(\theta) \top \widetilde{G}(\theta)$, where $\top$ is a triangular norm (or t-norm for short) [18]. The minimum is the largest t-norm. It is consistent with the definition of inclusion as $\widetilde{F} \subseteq \widetilde{G}$ iff $\widetilde{F} \leq \widetilde{G}$: $\widetilde{F} \wedge \widetilde{G}$ is then the largest fuzzy set included in $\widetilde{F}$ and $\widetilde{G}$. It is also idempotent, i.e., $\widetilde{F} \wedge \widetilde{F} = \widetilde{F}$; consequently, there is no reinforcement effect when two pieces of evidence are identical, which makes minimum-intersection combination useful for combining evidence from dependent and possibly redundant sources. In contrast, product intersection has a reinforcement effect that is appropriate when the sources are assumed to be independent [25, page 352].

The second difficulty when combining possibility distributions conjunctively is that the intersection of two fuzzy sets $\widetilde{F}$ and $\widetilde{G}$ may not be normal, and some normalization step

has to take place. The standard normalization procedure divides $\widetilde{F} \cap \widetilde{G}$ by its height (i.e., supremum). As a consequence, the associativity property is usually lost. However, the normalized product intersection, defined as

$$\widetilde{F} \odot \widetilde{G} := \frac{\widetilde{F} \cdot \widetilde{G}}{h(\widetilde{F} \cdot \widetilde{G})}, \tag{30}$$

is associative [25]. We give a simple proof of this well-known result below for completeness.

Proposition 3. *Let $\widetilde{F}$, $\widetilde{G}$ and $\widetilde{H}$ be fuzzy subsets of Θ, and let $\odot$ denote the normalized intersection based on the product t-norm. Then,*

$$(\widetilde{F} \odot \widetilde{G}) \odot \widetilde{H} = \widetilde{F} \odot (\widetilde{G} \odot \widetilde{H}).$$

Proof. The key property of the product-based intersection is that, for any $\alpha \geq 0$, $h((\alpha\widetilde{F}) \cdot \widetilde{G}) = h(\widetilde{F} \cdot (\alpha\widetilde{G})) = \alpha h(\widetilde{F} \cdot \widetilde{G})$. Using this property, we have

$$(\widetilde{F} \odot \widetilde{G}) \odot \widetilde{H} = \frac{\frac{\widetilde{F} \cdot \widetilde{G}}{h(\widetilde{F}\widetilde{G})} \cdot \widetilde{H}}{h\left(\frac{\widetilde{F} \cdot \widetilde{G}}{h(\widetilde{F} \cdot \widetilde{G})} \cdot \widetilde{H}\right)} = \frac{\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{H}/h(\widetilde{F} \cdot \widetilde{G})}{h(\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{H})/h(\widetilde{F} \cdot \widetilde{G})} = \frac{\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{H}}{h(\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{H})}.$$

and, by the commutativity of $\odot$,

$$\widetilde{F} \odot (\widetilde{G} \odot \widetilde{H}) = (\widetilde{G} \odot \widetilde{H}) \odot \widetilde{F} = \frac{\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{H}}{h(\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{H})}.$$

$\square$

The normalized intersection (30) is only defined if $h(\widetilde{F} \cdot \widetilde{G}) > 0$. A value of $h(\widetilde{F} \cdot \widetilde{G})$ close to zero signals conflicting evidence. Several authors have warned agains the use of normalized intersection in this case (see, e.g., [25, page 354]). Indeed, the division by a small number in (30) may result in high sensitivity of the combination to small changes in $\widetilde{F}$ or $\widetilde{G}$. Actually, conflict in the evidence may sometimes lead us to questioning its reliability. If we only assume that at least one source is reliable, we can only deduce that "$\boldsymbol{\theta}$ is $\widetilde{F} \cup \widetilde{G}$", where $\widetilde{F} \cup \widetilde{G}$ denotes the union of fuzzy sets $\widetilde{F}$ and $\widetilde{G}$. The union $\widetilde{F} \cup \widetilde{G}$ is usually defined as $(\widetilde{F} \cup_\perp \widetilde{G})(\theta) := \widetilde{F}(\theta) \perp \widetilde{G}(\theta)$, where $\perp$ is a t-conorm. To each t-norm $\top$ corresponds a dual t-conorm $\perp$ defined as $u \perp v = 1 - [(1-u)\top(1-v)]$. The t-conorms corresponding to the minimum and the product are, respectively, the maximum and the probabilistic sum $u \perp v = u + v - uv$.

Relation between possibility measures and belief functions. It can easily be shown that the mapping $N_{\widetilde{F}} : 2^\Omega \rightarrow [0,1]$ is completely monotone (11), i.e., it is a belief function, and $\Pi_{\widetilde{F}}$ is the dual plausibility function [24]. Actually, a belief function Bel_m and the associated plausibility function Pl_m verify the equalities

$$Bel_m(A \cap B) = Bel_m(A) \vee Bel_m(B)$$

11

and

$$Pl_m(A \cup B) = Pl_m(A) \vee Pl(_mB)$$

iff the corresponding mass function m is *consonant* [42], i.e., if for any pair of focal sets F_i and F_j, we have $F_i \subset F_j$ or $F_j \subset F_i$. To each possibility distribution $\pi_{\widetilde{F}}$ thus corresponds a unique consonant mass function $m_{\widetilde{F}}$, which can be recovered as follows [20]. Let $\theta_1, \ldots, \theta_q$ denote the elements of Θ, assumed to be indexed in such a way that

$$1 = \pi_{\widetilde{F}}(\theta_1) \geq \pi_{\widetilde{F}}(\theta_2) \geq \ldots \geq \pi(_{\widetilde{F}}\theta_q).$$

The corresponding mass function $m_{\widetilde{F}}$ is given by

$$m_{\widetilde{F}}(\{\theta_1, \ldots, \theta_i\}) = \pi_{\widetilde{F}}(\theta_i) - \pi_{\widetilde{F}}(\theta_{i+1}) \tag{31}$$

for $i = 1, \ldots, q - 1$, and $m_{\widetilde{F}}(\Theta) = \pi_{\widetilde{F}}(\theta_q)$. It can easily be checked that functions $Bel_{m_{\widetilde{F}}}$, $Pl_{m_{\widetilde{F}}}$, $pl_{m_{\widetilde{F}}}$ and $Q_{m_{\widetilde{F}}}$ are equal, respectively, to $N_{\widetilde{F}}$, $\Pi_{\widetilde{F}}$, $\pi_{\widetilde{F}}$ and $\Delta_{\widetilde{F}}$.

However, the combination of two consonant mass functions by Dempster's rule is no longer consonant. To combine two consonant belief functions Bel_1 and Bel_2 induced by independent sources, we must, therefore, consider the evidence on which they are based:

- If they are based on fully reliable but vague (fuzzy) evidence such as $\boldsymbol{\theta}$ is $\widetilde{F}$ and $\boldsymbol{\theta}$ is $\widetilde{G}$, then they should be combined by a conjunctive operator of possibility theory; the normalized product intersection (30) seems to be a good choice as it is associative;

- If they are based on uncertain but crisp (nonfuzzy) evidence pointing to consonant focal sets, then the corresponding consonant mass functions should be combined by Dempster's rule (14).

The two combination mechanisms yield different results but, as a consequence of Proposition 1, the contour functions are proportional, as illustrated by the following example.

Example 1. *Let $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$. Assume that we receive the following pieces of evidence from two independent and reliable sources:*

- *First piece of evidence: "$\boldsymbol{\theta}$ is $\widetilde{F}$", with*

$$\widetilde{F} = \left\{ \frac{\theta_1}{0.5}, \frac{\theta_2}{1}, \frac{\theta_3}{0.8}, \frac{\theta_4}{0.3} \right\};$$

- *Second piece of evidence: "$\boldsymbol{\theta}$ is $\widetilde{G}$", with*

$$\widetilde{G} = \left\{ \frac{\theta_1}{0.3}, \frac{\theta_2}{0.7}, \frac{\theta_3}{1}, \frac{\theta_4}{0.2} \right\};$$

These pieces of evidence can be represented by possibility distributions $\pi_{\widetilde{F}}$ and $\pi_{\widetilde{G}}$ and combined by (30), resulting in the combined possibility distribution $\pi_{\widetilde{F}\odot\widetilde{G}}$, with $\widetilde{F}\odot\widetilde{G}$ equal to

$$\widetilde{F}\odot\widetilde{G} = \left\{\frac{\theta_1}{0.15/0.8}, \frac{\theta_2}{0.7/0.8}, \frac{\theta_3}{1}, \frac{\theta_4}{0.06/0.8}\right\} = \left\{\frac{\theta_1}{0.1875}, \frac{\theta_2}{0.875}, \frac{\theta_3}{1}, \frac{\theta_4}{0.075}\right\}.$$

The consonant mass function $m_{\widetilde{F}\odot\widetilde{G}}$ corresponding to $\pi_{\widetilde{F}\odot\widetilde{G}}$ is

$$m_{\widetilde{F}\odot\widetilde{G}}(\{\theta_3\}) = 1 - 0.875 = 0.125$$
$$m_{\widetilde{F}\odot\widetilde{G}}(\{\theta_2, \theta_3\}) = 0.875 - 0.1875 = 0.6875$$
$$m_{\widetilde{F}\odot\widetilde{G}}(\{\theta_1, \theta_2, \theta_3\}) = 0.1875 - 0.075 = 0.1125$$
$$m_{\widetilde{F}\odot\widetilde{G}}(\Theta) = 0.075.$$

Consider now another situation in which we receive two pieces of evidence represented by the following consonant mass functions:

$$m_{\widetilde{F}}(\{\theta_2\}) = 1 - 0.8 = 0.2$$
$$m_{\widetilde{F}}(\{\theta_2, \theta_3\}) = 0.8 - 0.5 = 0.3$$
$$m_{\widetilde{F}}(\{\theta_1, \theta_2, \theta_3\}) = 0.5 - 0.3 = 0.2$$
$$m_{\widetilde{F}}(\Theta) = 0.3,$$

and

$$m_{\widetilde{G}}(\{\theta_3\}) = 1 - 0.7 = 0.3$$
$$m_{\widetilde{G}}(\{\theta_2, \theta_3\}) = 0.7 - 0.3 = 0.4$$
$$m_{\widetilde{G}}(\{\theta_1, \theta_2, \theta_3\}) = 0.3 - 0.2 = 0.1$$
$$m_{\widetilde{G}}(\Theta) = 0.2.$$

Mass functions $m_{\widetilde{F}}$ and $m_{\widetilde{G}}$ induce the same belief functions as, respectively, $\widetilde{F}$ and $\widetilde{G}$. Yet, they are combined differently by Dempster's rule. Their orthogonal sum $m_{\widetilde{F}} \oplus m_{\widetilde{G}}$ is

$$(m_{\widetilde{F}} \oplus m_{\widetilde{G}})(\{\theta_3\}) = 0.24/(1 - 0.06) \approx 0.255$$
$$(m_{\widetilde{F}} \oplus m_{\widetilde{G}})(\{\theta_2\}) = 0.14/(1 - 0.06) \approx 0.149$$
$$(m_{\widetilde{F}} \oplus m_{\widetilde{G}})(\{\theta_2, \theta_3\}) = 0.41/(1 - 0.06) \approx 0.436$$
$$(m_{\widetilde{F}} \oplus m_{\widetilde{G}})(\{\theta_1, \theta_2, \theta_3\}) = 0.09/(1 - 0.06) \approx 0.0957$$
$$(m_{\widetilde{F}} \oplus m_{\widetilde{G}})(\Theta) = 0.06/(1 - 0.06) \approx 0.0638,$$

which is different from $m_{\widetilde{F}\odot\widetilde{G}}$. In particular, $m_{\widetilde{F}}\oplus m_{\widetilde{G}}$ is not consonant. Its contour function of $m_{\widetilde{F}} \oplus m_{\widetilde{G}}$ is

$$pl_{m_{\widetilde{F}}\oplus m_{\widetilde{G}}}(\theta_1) = (0.09 + 0.06)/(1 - 0.06) \approx 0.160$$
$$pl_{m_{\widetilde{F}}\oplus m_{\widetilde{G}}}(\theta_2) = (0.14 + 0.41 + 0.09 + 0.06)/(1 - 0.06) \approx 0.745$$
$$pl_{m_{\widetilde{F}}\oplus m_{\widetilde{G}}}(\theta_3) = (0.24 + 0.41 + 0.09 + 0.06)/(1 - 0.06) \approx 0.851$$
$$pl_{m_{\widetilde{F}}\oplus m_{\widetilde{G}}}(\theta_4) = 0.06/(1 - 0.06) \approx 0.0638.$$

It can be checked that $\pi_{\widetilde{F}\odot\widetilde{G}}$ and $pl_{m_{\widetilde{F}}\oplus m_{\widetilde{G}}}$ are proportional, with $\pi_{\widetilde{F}\odot\widetilde{G}}/pl_{m_{\widetilde{F}}\oplus m_{\widetilde{G}}} = 1.175$.

The above considerations show that it is important, in practice, to determine whether a piece of evidence should be represented by a possibility distribution or by a consonant mass function. It seems reasonable to use the former representation for reliable but fuzzy evidence such as conveyed, e.g., by natural language, as in the proposition "John is tall". More surprisingly, it appears that *statistical evidence* should also be represented in that way, as will be shown in Section 4. In contrast, consonant evidence usually arises when combining several elementary pieces of evidence, such as simple mass functions of the form $m_i(F_i) = p_i$, $m_i(\Theta) = 1 - p_i$ with $F_1 \subseteq \ldots \subseteq F_f$. Such simple mass functions may be obtained by combining expert opinions or sensor readings telling us that $F_i \subseteq \Theta$ with confidence degree p_i.

Possibility and necessity of a fuzzy event. In [57], Zadeh defines the possibility and the necessity of a fuzzy event $\widetilde{A} \in \mathcal{F}(\Theta)$ as

$$\Pi_{\widetilde{F}}^{(S)}(\widetilde{A}) := \max_{\theta \in \Theta} (\widetilde{A} \wedge \widetilde{F})(\theta) = h(\widetilde{A} \wedge \widetilde{F}). \tag{32}$$

and

$$N_{\widetilde{F}}^{(S)}(\widetilde{A}) := 1 - \Pi_{\widetilde{F}}^{(S)}(\widetilde{A}^c) \tag{33a}$$

$$= 1 - \max_{\theta \in \Theta} [1 - A(\theta)] \wedge \widetilde{F}(\theta) \tag{33b}$$

$$= \min_{\theta \in \Theta} \left\{ 1 - [1 - \widetilde{A}(\theta)] \wedge \widetilde{F}(\theta) \right\} \tag{33c}$$

$$= \min_{\theta \in \Theta} \widetilde{A}(\theta) \vee [1 - \widetilde{F}(\theta)] \tag{33d}$$

$$= \min_{\theta \in \Theta} (\widetilde{A} \vee \widetilde{F}^c)(\theta). \tag{33e}$$

We note that $\Pi_{\widetilde{F}}^{(S)}(\widetilde{A})$ and $N_{\widetilde{F}}^{(S)}(\widetilde{A})$ are Sugeno integrals of the mapping $\widetilde{A} : \Theta \to [0, 1]$ with respect to $\Pi_{\widetilde{F}}$ and $N_{\widetilde{F}}$, respectively [18, Section 7.6.2]. We can also remark that $\mathrm{Int}(\widetilde{A}, \widetilde{F}) := h(\widetilde{A} \wedge \widetilde{F})$ can be seen as a *degree of intersection* of $\widetilde{A}$ and $\widetilde{F}$, whereas $\mathrm{Incl}(\widetilde{F}, \widetilde{A}) := \min_{\theta \in \Theta}(\widetilde{A} \vee \widetilde{F}^c)(\theta)$ can be seen as a *degree of inclusion* of $\widetilde{F}$ in $\widetilde{A}$.

It is easy to check that $\Pi_{\widetilde{F}}^{(S)}$ and $N_{\widetilde{F}}^{(S)}$ are still, respectively, possibility and necessity measures in the lattice $(\mathcal{F}(\Theta), \wedge, \vee)$, as

$$\Pi_{\widetilde{F}}^{(S)}(\widetilde{A} \vee \widetilde{B}) = \Pi_{\widetilde{F}}^{(S)}(\widetilde{A}) \vee \Pi_{\widetilde{F}}^{(S)}(\widetilde{B}) \tag{34a}$$

and

$$N_{\widetilde{F}}^{(S)}(\widetilde{A} \wedge \widetilde{B}) = N_{\widetilde{F}}^{(S)}(\widetilde{A}) \wedge N_{\widetilde{F}}^{(S)}(\widetilde{B}) \tag{34b}$$

for all fuzzy sets $\widetilde{A}$ and $\widetilde{B}$, which generalize (4)-(5). It is also easy to show that $N_{\widetilde{F}}^{(S)}(\widetilde{A}) \leq \Pi_{\widetilde{F}}^{(S)}(\widetilde{A})$ for all $\widetilde{A} \in \mathcal{F}(\Theta)$ [21]. Dubois and Prade [21] also show that $N_{\widetilde{F}}^{(S)}$ is a belief function in the lattice $(\mathcal{F}(\Theta), \wedge, \vee)$, and $\Pi_{\widetilde{F}}^{(S)}$ is the dual plausibility function.

Dubois *et al.* review generalizations of (32)-(33) as well as alternative definitions in [18, Section 7.6]. In particular, they remark that we can identify $\Pi_{\widetilde{F}}$ and $N_{\widetilde{F}}$ to, respectively, the plausibility function $Pl_{m_{\widetilde{F}}}$ and the belief function $Bel_{m_{\widetilde{F}}}$ induced by the consonant mass function $m_{\widetilde{F}}$ defined by (31). We can then define the possibility and necessity of a fuzzy event $\widetilde{A}$ from (18) and (19) by the Choquet integrals of the mapping $\widetilde{A} : \Theta \to [0, 1]$ with respect to the non-additive functions $\Pi_{\widetilde{F}}$ and $N_{\widetilde{F}}$:

$$\Pi_{\widetilde{F}}^{(C)}(\widetilde{A}) := \int_0^1 \Pi_{\widetilde{F}}({}^\alpha \widetilde{A}) d\alpha = \sum_{A \subseteq \Theta} m_{\widetilde{F}}(A) \max_{\theta \in A} \widetilde{A}(\theta) = Pl_{m_{\widetilde{F}}}(\widetilde{A}) \tag{35a}$$

$$N_{\widetilde{F}}^{(C)}(\widetilde{A}) := \int_0^1 N_{\widetilde{F}}({}^\alpha \widetilde{A}) d\alpha = \sum_{A \subseteq \Theta} m_{\widetilde{F}}(A) \min_{\theta \in A} \widetilde{A}(\theta) = Bel_{m_{\widetilde{F}}}(\widetilde{A}). \tag{35b}$$

From (25), we then have

$$\Pi_{\widetilde{F}}^{(C)}(\widetilde{A}) = \int_0^1 \left(\int_0^1 \Pi_{\beta \widetilde{F}}({}^\alpha \widetilde{A}) d\beta \right) d\alpha = \int_0^1 \Pi_{\beta \widetilde{F}}(\widetilde{A}) d\beta \tag{36a}$$

and

$$N_{\widetilde{F}}^{(C)}(\widetilde{A}) = \int_0^1 \left(\int_0^1 N_{\beta \widetilde{F}}({}^\alpha \widetilde{A}) d\beta \right) d\alpha = \int_0^1 N_{\beta \widetilde{F}}(\widetilde{A}) d\beta, \tag{36b}$$

with

$$\Pi_{\beta \widetilde{F}}(\widetilde{A}) = \int_0^1 \Pi_{\beta \widetilde{F}}({}^\alpha \widetilde{A}) d\alpha = \max_{\theta \in \beta \widetilde{F}} \widetilde{A}(\theta) = \max_{\{\theta : \widetilde{F}(\theta) \geq \beta\}} \widetilde{A}(\theta) \tag{36c}$$

and

$$N_{\beta \widetilde{F}}(\widetilde{A}) = \int_0^1 N_{\beta \widetilde{F}}({}^\alpha \widetilde{A}) d\alpha = \min_{\theta \in \beta \widetilde{F}} \widetilde{A}(\theta) = \min_{\{\theta : \widetilde{F}(\theta) \geq \beta\}} \widetilde{A}(\theta). \tag{36d}$$

As $N_{\widetilde{F}}^{(S)}$, function $N_{\widetilde{F}}^{(C)}$ defined by (35b) is a belief function on the lattice $(\mathcal{F}(\Theta), \wedge, \vee)$, and $\Pi_{\widetilde{F}}^{(C)}$ is its dual plausibility function. However, $\Pi_{\widetilde{F}}^{(C)}$ and $N_{\widetilde{F}}^{(C)}$ are no longer possibility and necessity measures, as they fail to satisfy the basic axioms (34) [21].

The following example illustrates the difference between $\Pi_{\widetilde{F}}^{(S)}$ and $\Pi_{\widetilde{F}}^{(C)}$.

Example 2. *Consider the fuzzy sets $\widetilde{F}$ and $\widetilde{G}$ of Example 1. We have*

$$\Pi_{\widetilde{F}}^{(S)}(\widetilde{G}) = h(\widetilde{F} \wedge \widetilde{G}) = h\left(\left\{ \frac{\theta_1}{0.3}, \frac{\theta_2}{0.7}, \frac{\theta_3}{0.8}, \frac{\theta_4}{0.2} \right\} \right) = 0.8$$

and

$$N_{\widetilde{F}}^{(S)}(\widetilde{G}) = 1 - h(\widetilde{F} \wedge \widetilde{G}^c) = 1 - h\left(\left\{ \frac{\theta_1}{0.5}, \frac{\theta_2}{0.3}, \frac{\theta_3}{0}, \frac{\theta_4}{0.3} \right\} \right) = 1 - 0.5 = 0.5,$$

but

$$\Pi_{\widetilde{F}}^{(C)}(\widetilde{G}) = 0.2 \times 0.7 + 0.3 \times 1 + 0.2 \times 1 + 0.3 \times 1 = 0.94.$$

and

$$N_{\widetilde{F}}^{(C)}(\widetilde{G}) = 0.2 \times 0.7 + 0.3 \times 0.7 + 0.2 \times 0.3 + 0.3 \times 0.2 = 0.47.$$

To conclude this section, we can remark that the guaranteed possibility function (27) can be extended to fuzzy events as well. It can easily be seen that, for crisp A,

$$\min_{\theta \in A} \widetilde{F}(\theta) = \min_{\theta \in \Theta} (A^c \vee \widetilde{F})(\theta) = \mathrm{Incl}(A, \widetilde{F}).$$

A natural definition for the guaranteed possibility of fuzzy event $\widetilde{A}$, in the spirit of Zadeh's definitions (32)-(33) for the possibility and belief of a fuzzy event is, thus,

$$\Delta_{\widetilde{F}}^{(S)}(\widetilde{A}) := \mathrm{Incl}(\widetilde{A}, \widetilde{F}) = \min_{\theta \in \Theta} (\widetilde{A}^c \vee \widetilde{F})(\theta). \tag{37}$$

As a generalization of (29), we have

$$\Delta_{\widetilde{F}}^{(S)}(\widetilde{A} \vee \widetilde{B}) = \Delta_{\widetilde{F}}^{(S)}(\widetilde{A}) \wedge \Delta_{\widetilde{F}}^{(S)}(\widetilde{B}). \tag{38}$$

Alternatively, we can define the guaranteed possibility of fuzzy event $\widetilde{A}$ from (22) by the Choquet integral

$$\Delta_{\widetilde{F}}^{(C)}(\widetilde{A}) := \int_0^1 \Delta_{\widetilde{F}}(^\alpha \widetilde{A}) d\alpha = \sum_{A \subseteq \Theta} m_{\widetilde{F}}(A) \left(1 - \max_{\theta \notin A} \widetilde{A}(\theta)\right) d\alpha = Q_{m_{\widetilde{F}}}(\widetilde{A}). \tag{39}$$

Unfortunately, equality (38) does not hold anymore with this alternative definition.

Example 3. *Consider again the fuzzy sets $\widetilde{F}$ and $\widetilde{G}$ of Examples 1 and 2. From*

$$\widetilde{G}^c \vee \widetilde{F} = \left\{\frac{\theta_1}{0.7}, \frac{\theta_2}{1}, \frac{\theta_3}{0.8}, \frac{\theta_4}{0.8}\right\},$$

we have $\Delta_{\widetilde{F}}^{(S)}(\widetilde{G}) = 0.7$, but

$$\Delta_{\widetilde{F}}^{(C)}(\widetilde{G}) = 0.2(1 - 1) + 0.3(1 - 0.3) + 0.2(1 - 0.2) + 0.3 = 0.67.$$

Now, let

$$\widetilde{H} = \left\{\frac{\theta_1}{1}, \frac{\theta_2}{0.6}, \frac{\theta_3}{0.}, \frac{\theta_4}{0.1}\right\}.$$

We have

$$\widetilde{H}^c \vee \widetilde{F} = \left\{\frac{\theta_1}{0.5}, \frac{\theta_2}{1}, \frac{\theta_3}{0.8}, \frac{\theta_4}{0.9}\right\},$$

hence $\Delta_{\widetilde{F}}^{(S)}(\widetilde{H}) = 0.5$, and

$$(\widetilde{G} \vee \widetilde{H})^c \vee \widetilde{F} = \left\{\frac{\theta_1}{0.5}, \frac{\theta_2}{1}, \frac{\theta_3}{0.8}, \frac{\theta_4}{0.8}\right\},$$

hence $\Delta_{\widetilde{F}}^{(S)}(\widetilde{G} \vee \widetilde{H}) = 0.5 = \Delta_{\widetilde{F}}^{(S)}(\widetilde{G}) \wedge \Delta_{\widetilde{F}}^{(S)}(\widetilde{H})$. But we have

$$\Delta_{\widetilde{F}}^{(C)}(\widetilde{H}) = 0.2(1 - 1) + 0.3(1 - 1) + 0.2(1 - 0.1) + 0.3 = 0.48$$

and

$$\Delta_{\widetilde{F}}^{(C)}(\widetilde{G} \vee \widetilde{H}) = 0.2(1 - 1) + 0.3(1 - 1) + 0.2(1 - 0.2) + 0.3 = 0.46$$
$$\neq \Delta_{\widetilde{F}}^{(C)}(\widetilde{G}) \wedge \Delta_{\widetilde{F}}^{(C)}(\widetilde{H}).$$

16

3. Uncertain and fuzzy information: fuzzy mass functions

To handle cases where one consonant belief function is based on fuzzy evidence and the other one on uncertain evidence, or to handle evidence that is *both* uncertain and fuzzy, we need to generalize DS and possibility theories. Such a generalization is exposed in this section. We first review the notion of fuzzy mass function (Section 3.1). In Section 3.2, we propose a combination operator that extends both Dempster's rule (14) and the normalized product (30); we also define extensions of the disjunctive operators reviewed in Section 2.2 and 2.3. The degrees of belief and plausibility of fuzzy events are then defined in Section 3.3.

3.1. Fuzzy mass functions

Following [58], let us assume that we receive uncertain and fuzzy information, which can be modeled as follows. As in Section 2.2, we assume that the evidence can be interpreted in different ways, and the set of interpretations is denoted by Ω. If interpretation $\omega \in \Omega$ holds, then we know for sure that proposition "$\boldsymbol{\theta}$ is $\widetilde{\Gamma}(\omega)$" is true, where $\widetilde{\Gamma}(\omega)$ is a normal fuzzy subset of Θ. Denoting by $\mathcal{F}^*(\Theta)$ the set of all normal fuzzy subsets of Θ, we thus have a mapping $\widetilde{\Gamma} : \Omega \to \mathcal{F}^*(\Theta)$. If, as before, we assume the existence of a probability measure P on $(\Omega, 2^\Omega)$, then the tuple $(\Omega, 2^\Omega, P, \widetilde{\Gamma})$ is a *random fuzzy set* [26, 33, 34] (also called a "fuzzy random variable" when Θ is $\mathbb{R}^p$).

Let $\widetilde{m}$ be the mapping from $\mathcal{F}^*(\Theta)$ to $[0,1]$ defined as

$$\widetilde{m}(\widetilde{F}) := P(\{\omega \in \Omega : \widetilde{\Gamma}(\omega) = \widetilde{F}\}).$$

Because Ω is assumed to be finite, there is only a finite number of fuzzy subsets $\widetilde{F}$ such that $\widetilde{m}(\widetilde{F}) > 0$, called the (fuzzy) focal sets of m. The set of focal sets is denoted as $\mathbb{F}(m) = \{\widetilde{F}_1, \ldots, \widetilde{F}_f\}$. We also use the notation $m_i := m(\widetilde{F}_i)$, $i = 1, \ldots, n$. Mapping $\widetilde{m}$ is called a *fuzzy mass function*[1]. The notion of fuzzy mass function extends that of DS mass function recalled in Section 2.2. The number $\widetilde{m}(\widetilde{F}_i)$ is interpreted as the degree with which the evidence supports the proposition $\boldsymbol{\theta}$ is $\widetilde{F}_i$, without supporting any more specific proposition.

If interpretation ω holds, we know that "$\boldsymbol{\theta}$ is $\widetilde{\Gamma}(\omega)$". The possibility and necessity of a subset $A \subseteq \Theta$ are, respectively, $\Pi_{\widetilde{\Gamma}(\omega)}(A)$ and $N_{\widetilde{\Gamma}(\omega)}(A)$ defined by (23) and (24). As we only know that interpretation ω holds with probability $P(\{\omega\})$, we can compute the expected possibility and the expected necessity [58] as

$$Pl_{\widetilde{m}}(A) = \sum_{\omega \in \Omega} P(\{\omega\}) \Pi_{\widetilde{\Gamma}(\omega)}(A) = \sum_{i=1}^{f} m_i \Pi_{\widetilde{F}_i}(A) = \sum_{i=1}^{f} m_i \max_{\theta \in A} \widetilde{F}_i(\theta) \tag{40}$$

---

[1]This notion should not be confused with that of *fuzzy-valued mass function* introduced in [10]. A fuzz-valued mass function assigns fuzzy numbers to crisp focal sets, and can be interpreted as a fuzzy set of crisp mass functions. We could, of course, "fuzzify" both the masses and the focal sets; such a generalization will not be considered in this paper.

and

$$Bel_{\widetilde{m}}(A) = \sum_{\omega \in \Omega} P(\{\omega\}) N_{\widetilde{\Gamma}(\omega)}(A) = \sum_{i=1}^{f} m_i N_{\widetilde{F}_i}(A) = \sum_{i=1}^{f} m_i \min_{\theta \notin A}[1 - \widetilde{F}_i(\theta)]. \qquad (41)$$

Functions $Pl_{\widetilde{m}}$ and $Bel_{\widetilde{m}}$ are, respectively, mixtures of possibility and necessity measures. As we have seen that each necessity measure $N_{\widetilde{F}_i}$ is a belief function, and the set of belief functions is convex, $Bel_{\widetilde{m}}$ is still a belief function, and $Pl_{\widetilde{m}}$ (which verifies $Pl_{\widetilde{m}}(A) = 1 - Bel_{\widetilde{m}}(A^c)$ for all $A \subseteq \Theta$) is the dual plausibility function. The contour function of $\widetilde{m}$ is equal to the mean of the membership functions of its focal sets:

$$pl_{\widetilde{m}}(\theta) = \sum_{i=1}^{f} m_i \widetilde{F}_i(\theta). \qquad (42)$$

We can also define the commonality of A as its expected guaranteed possibility from (27):

$$Q_{\widetilde{m}}(A) = \sum_{\omega \in \Omega} P(\{\omega\}) \Delta_{\widetilde{\Gamma}(\omega)}(A) = \sum_{i=1}^{f} m_i \Delta_{\widetilde{F}_i}(A) = \sum_{i=1}^{f} m_i \min_{\theta \in A} \widetilde{F}_i(\theta). \qquad (43)$$

A fuzzy mass function $\widetilde{m}$ with focal sets $\widetilde{F}_1, \ldots, \widetilde{F}_f$ and masses $m_1, \ldots, m_f$ can also be described as a collection of standard (crisp) mass functions ${}^{\alpha}\widetilde{m}$ with focal sets ${}^{\alpha}\widetilde{F}_1, \ldots, {}^{\alpha}\widetilde{F}_f$ and the same masses $m_1, \ldots, m_f$. Each mass function ${}^{\alpha}\widetilde{m}$ is induced by the random set $(\Omega, 2^{\Omega}, P, {}^{\alpha}\widetilde{\Gamma})$, with ${}^{\alpha}\widetilde{\Gamma}(\omega) = {}^{\alpha}[\widetilde{\Gamma}(\omega)]$. From (25) and (27), we have

$$Bel_{\widetilde{m}}(A) = \int_0^1 Bel_{{}^{\alpha}\widetilde{m}}(A) d\alpha, \qquad (44a)$$

$$Pl_{\widetilde{m}}(A) = \int_0^1 Pl_{{}^{\alpha}\widetilde{m}}(A) d\alpha, \qquad (44b)$$

and

$$Q_{\widetilde{m}}(A) = \int_0^1 Q_{{}^{\alpha}\widetilde{m}}(A) d\alpha \qquad (44c)$$

for all $A \subseteq \Theta$. We note that equalities (44a) and (44b) are proved in [5] in a more general setting.

It is important to remark that the concept of fuzzy mass functions allows us to generalize both DS theory and possibility theory. In particular, a *logical* fuzzy mass function $\widetilde{m}$ with a single fuzzy focal set $\widetilde{F}$ is equivalent to a possibility distribution $\pi = \widetilde{F}$.

Remark 1. *In "classical" DS theory there is a one-to-one correspondence between (crisp) mass functions and belief functions. The fundamental reason is that, for a given belief function Bel, the system of linear equations $Bel(A) = \sum_{B \subseteq A} m(B)$ for all $\emptyset \neq A \subseteq \Omega$ has a unique solution. In the "fuzzified" version of DS theory, there is a many-to-one correspondence between fuzzy mass functions and belief functions, i.e., a given belief function can be obtained from many fuzzy mass functions. The simplest example is that of a belief*

function verifying $Bel(A \cap B) = Bel(A) \wedge Bel(B)$ for all $A, B \subseteq \Theta$ (i.e., a necessity measure), which, as shown in Section 2.3, can be induced both by a crisp consonant mass function and by a logical fuzzy mass function.

In the following section, we define a combination operator for fuzzy mass functions that extends both the normalized product of possibility distributions (30) and Dempster's rule (14).

3.2. Combination of fuzzy mass functions

Let us now assume that we have two fuzzy mass functions $\widetilde{m}_1$ and $\widetilde{m}_2$ induced by independent random fuzzy sets $(\Omega_1, 2^{\Omega_1}, P_1, \widetilde{\Gamma}_1)$ and $(\Omega_2, 2^{\Omega_2}, P_2, \widetilde{\Gamma}_2)$. If interpretations $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ both hold, then we can infer the fuzzy proposition "$\boldsymbol{\theta}$ is $\widetilde{\Gamma}_{\cap_\top}(\omega_1, \omega_2)$", with $\widetilde{\Gamma}_{\cap_\top}(\omega_1, \omega_2) := \widetilde{\Gamma}_1(\omega_1) \cap_\top \widetilde{\Gamma}_2(\omega_2)$, where $\cap_\top$ is a fuzzy intersection operator based on a t-norm $\top$. However, the random fuzzy set $(\Omega_1 \times \Omega_2, 2^{\Omega_1 \times \Omega_2}, P_1 \otimes P_2, \widetilde{\Gamma}_{\cap_\top})$ may not be normalized, as some fuzzy sets $\widetilde{\Gamma}_{\cap_\top}(\omega_1, \omega_2)$ may not be normal. To obtain a normal random fuzzy set, we can replace $\widetilde{\Gamma}_{\cap_\top}$ by the mapping

$$\widetilde{\Gamma}_{\overline{\cap}_\top}(\omega_1, \omega_2) := \widetilde{\Gamma}_1(\omega_1) \overline{\cap}_\top \widetilde{\Gamma}_2(\omega_2) = \frac{\widetilde{\Gamma}_1(\omega_1) \cap_\top \widetilde{\Gamma}_2(\omega_2)}{h(\widetilde{\Gamma}_1(\omega_1) \cap_\top \widetilde{\Gamma}_2(\omega_2))}, \tag{45}$$

where $\overline{\cap}_\top$ denotes normalized $\top$-intersection, and we can condition $P_1 \otimes P_2$ by the following fuzzy subset $\widetilde{\Theta}^*$ of $\Omega_1 \times \Omega_2$, which is a natural generalization of Θ^* in (13):

$$\widetilde{\Theta}^*(\omega_1, \omega_2) = h(\widetilde{\Gamma}_{\cap_\top}(\omega_1, \omega_2)), \quad \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2. \tag{46}$$

Using Zadeh's definition of the probability of a fuzzy event (16), the conditional probability $(P_1 \otimes P_2)(\cdot \mid \widetilde{\Theta}^*)$ is

$$(P_1 \otimes P_2)(B \mid \widetilde{\Theta}^*) = \frac{(P_1 \otimes P_2)(B \cap \widetilde{\Theta}^*)}{(P_1 \otimes P_2)(\widetilde{\Theta}^*)}$$
$$= \frac{\sum_{(\omega_1, \omega_2) \in B} P_1(\omega_1) P_2(\omega_2) h(\widetilde{\Gamma}_{\cap_\top}(\omega_1, \omega_2))}{\sum_{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2} P_1(\omega_1) P_2(\omega_2) h(\widetilde{\Gamma}_{\cap_\top}(\omega_1, \omega_2))},$$

for all $B \subseteq \Omega_1 \times \Omega_2$. The fuzzy mass function generated by the random fuzzy set $(\Omega_1 \times \Omega_2, 2^{\Omega_1 \times \Omega_2}, (P_1 \otimes P_2)(\cdot \mid \widetilde{\Theta}^*), \widetilde{\Gamma}'_{\cap_\top})$ is then

$$(\widetilde{m}_1 \,\widecheck{\cap}_\top\, \widetilde{m}_2)(\widetilde{F}) := \frac{\sum_{\widetilde{G} \overline{\cap}_\top \widetilde{H} = \widetilde{F}} h(\widetilde{G} \cap_\top \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H})}{\sum_{(\widetilde{G}, \widetilde{H}) \in \mathbb{F}(\widetilde{m}_1) \times \mathbb{F}(\widetilde{m}_2)} h(\widetilde{G} \cap_\top \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H})} \tag{47}$$

for all $\widetilde{F} \in \mathcal{F}^*(\Theta)$. This operation was proposed by Yen [53], using $\top = \min$. The normalization operation in (47) was called "soft normalization" by Yager [52].

In general, the operation $\widecheck{\cap}_\top$ defined by (47) is not associative. However, it is associative if we use the *product t-norm* for defining the fuzzy set intersection. In the following, we will

thus define the *orthogonal sum* $\widetilde{m}_1 \oplus \widetilde{m}_2$ of two fuzzy mass functions $\widetilde{m}_1$ and $\widetilde{m}_2$ (where the symbol $\oplus$ has the same meaning as $\text{\textcircled{m}}_\top$ with $\top = $ product) as

$$(\widetilde{m}_1 \oplus \widetilde{m}_2)(\widetilde{F}) := \frac{\sum_{\widetilde{G} \odot \widetilde{H} = \widetilde{F}} h(\widetilde{G} \cdot \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H})}{\sum_{(\widetilde{G}, \widetilde{H}) \in \mathbb{F}(\widetilde{m}_1) \times \mathbb{F}(\widetilde{m}_2)} h(\widetilde{G} \cdot \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H})}, \tag{48}$$

for all $\widetilde{F} \in \mathcal{F}^*(\Theta)$. The operation defined by (48) will be called the *generalized product-intersection* rule. The denominator of the right-hand side of (48) can be denoted as $1 - \kappa$, where κ can be interpreted as the *degree of conflict* between fuzzy mass functions $\widetilde{m}_1$ and $\widetilde{m}_2$. It is clear that (48) is a proper generalization of (14), i.e., it gives the same result when the focal sets of $\widetilde{m}_1$ and $\widetilde{m}_2$ are crisp. It also boils down to normalized intersection (30) when combining two logical fuzzy mass functions $\widetilde{m}_1$ and $\widetilde{m}_2$ such that $\widetilde{m}_1(\widetilde{F}) = 1$ and $\widetilde{m}_2(\widetilde{G}) = 1$. As a consequence of Proposition 3, the generalized product-intersection rule is associative, as stated in the following proposition.

Proposition 4. *Let m_1, m_2 and m_3 be fuzzy mass functions of Θ, and let $\oplus$ denote the generalized product-intersection defined by (48). Then,*

$$(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3).$$

Proof. We have

$$[(m_1 \oplus m_2) \oplus m_3](\widetilde{A}) = \frac{\sum_{\widetilde{H} \odot \widetilde{K} = \widetilde{A}} (m_1 \oplus m_2)(\widetilde{H}) m_3(\widetilde{K}) h(\widetilde{H} \cdot \widetilde{K})}{\sum_{\widetilde{H}, \widetilde{K}} (m_1 \oplus m_2)(\widetilde{H}) m_3(\widetilde{K}) h(\widetilde{H} \cdot \widetilde{K})}$$

$$= \frac{\sum_{\widetilde{H} \odot \widetilde{K} = \widetilde{A}} \left[\frac{\sum_{\widetilde{F} \odot \widetilde{G} = \widetilde{H}} m_1(\widetilde{F}) m_2(\widetilde{G}) h(\widetilde{F} \cdot \widetilde{G})}{\sum_{\widetilde{F}, \widetilde{G}} m_1(\widetilde{F}) m_2(\widetilde{G})} \right] m_3(\widetilde{K}) h(\widetilde{H} \cdot \widetilde{K})}{\sum_{\widetilde{H}, \widetilde{K}} \left[\frac{\sum_{\widetilde{F} \odot \widetilde{G} = \widetilde{H}} m_1(\widetilde{F}) m_2(\widetilde{G}) h(\widetilde{F} \cdot \widetilde{G})}{\sum_{\widetilde{F}, \widetilde{G}} m_1(\widetilde{F}) m_2(\widetilde{G})} \right] m_3(\widetilde{K}) h(\widetilde{H} \widetilde{K})}$$

$$= \frac{\sum_{\widetilde{F} \odot \widetilde{G} \odot \widetilde{K} = \widetilde{A}} m_1(\widetilde{F}) m_2(\widetilde{G}) m_3(\widetilde{H}) h(\widetilde{F} \cdot \widetilde{G}) h((\widetilde{F} \odot \widetilde{G}) \cdot \widetilde{K})}{\sum_{\widetilde{F}, \widetilde{G}, \widetilde{K}} m_1(\widetilde{F}) m_2(\widetilde{G}) m_3(\widetilde{H}) h(\widetilde{F} \cdot \widetilde{G}) h((\widetilde{F} \odot \widetilde{G}) \cdot \widetilde{K})}.$$

Now,

$$h(\widetilde{F} \cdot \widetilde{G}) h((\widetilde{F} \odot \widetilde{G}) \cdot \widetilde{K}) = h(\widetilde{F} \cdot \widetilde{G}) h\left(\frac{\widetilde{F} \cdot \widetilde{G}}{h(\widetilde{F} \cdot \widetilde{G})} \cdot \widetilde{K} \right) = h(\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{K}).$$

Hence,

$$[(m_1 \oplus m_2) \oplus m_3](\widetilde{A}) = \frac{\sum_{\widetilde{F} \odot \widetilde{G} \odot \widetilde{K} = \widetilde{A}} m_1(\widetilde{F}) m_2(\widetilde{G}) m_3(\widetilde{H}) h(\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{K})}{\sum_{\widetilde{F}, \widetilde{G}, \widetilde{K}} m_1(\widetilde{F}) m_2(\widetilde{G}) m_3(\widetilde{H}) h(\widetilde{F} \cdot \widetilde{G} \cdot \widetilde{K})}. \tag{49}$$

Consequently, we can permute the indices in (49) and write

$$(m_1 \oplus m_2) \oplus m_3 = (m_2 \oplus m_3) \oplus m_1 = m_1 \oplus (m_2 \oplus m_3).$$

$\square$

Example 4. *Continuing Example 1, assume that we have the following two fuzzy mass functions:*

$$m_1(\widetilde{F}) := 0.6, \quad m_1(\Theta) := 0.4$$

and

$$m_2(\widetilde{G}) := 0.7, \quad m_2(\Theta) := 0.3,$$

where $\widetilde{F}$ and $\widetilde{G}$ are defined as in Example 1. As $h(\widetilde{F} \cdot \widetilde{G}) = 0.8$, we have

$$(m_1 \oplus m_2)(\widetilde{F} \odot \widetilde{G}) = \frac{0.6 \times 0.7 \times 0.8}{0.6 \times 0.7 \times 0.8 + 0.6 \times 0.3 + 0.4 \times 0.7 + 0.4 \times 0.3} \approx 0.37$$

$$(m_1 \oplus m_2)(\widetilde{F}) = \frac{0.6 \times 0.3}{0.6 \times 0.7 \times 0.8 + 0.6 \times 0.3 + 0.4 \times 0.7 + 0.4 \times 0.3} \approx 0.20$$

$$(m_1 \oplus m_2)(\widetilde{G}) = \frac{0.7 \times 0.4}{0.6 \times 0.7 \times 0.8 + 0.6 \times 0.3 + 0.4 \times 0.7 + 0.4 \times 0.3} \approx 0.31$$

$$(m_1 \oplus m_2)(\Theta) = \frac{0.3 \times 0.4}{0.6 \times 0.7 \times 0.8 + 0.6 \times 0.3 + 0.4 \times 0.7 + 0.4 \times .0.3} \approx 0.13.$$

The following propositions generalize Propositions 1 and 2.

Proposition 5. *Let $\widetilde{m}_1$ and $\widetilde{m}_2$ be two fuzzy mass functions with contour functions $pl_{\widetilde{m}_1}$ and $pl_{\widetilde{m}_2}$. Then, the contour function of $\widetilde{m}_1 \oplus \widetilde{m}_2$ is given by*

$$pl_{\widetilde{m}_1 \oplus \widetilde{m}_2} = \frac{pl_{\widetilde{m}_1} pl_{\widetilde{m}_2}}{1 - \kappa}$$

with

$$\kappa = 1 - \sum_{(\widetilde{G}, \widetilde{H}) \in \mathbb{F}(\widetilde{m}_1) \times \mathbb{F}(\widetilde{m}_2)} h(\widetilde{G} \cdot \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H}).$$

Proof. Let $\widetilde{F}_1, \ldots, \widetilde{F}_{f_1}$ and $\widetilde{G}_1, \ldots, \widetilde{G}_{f_2}$ be the focal sets of $\widetilde{m}_1$ and $\widetilde{m}_2$, respectively, with corresponding masses $m_{11}, \ldots, m_{1f_1}$ and $m_{21}, \ldots, m_{2f_2}$. From (42),

$$pl_{\widetilde{m}_1 \oplus \widetilde{m}_2}(\theta) = \sum_{i,j} \frac{m_{1i} m_{2j} h(\widetilde{F}_i \widetilde{G}_j)}{1 - \kappa} (\widetilde{F}_i \odot \widetilde{G}_j)(\theta)$$

$$= \frac{\sum_{i,j} m_{1i} m_{2j} \widetilde{F}_i(\theta) \widetilde{G}_j(\theta)}{1 - \kappa}$$

$$= \frac{\left(\sum_i m_{1i} \widetilde{F}_i(\theta)\right) \left(\sum_j m_{2j} \widetilde{G}_j(\theta)\right)}{1 - \kappa} = \frac{pl_{\widetilde{m}_1}(\theta) pl_{\widetilde{m}_2}(\theta)}{1 - \kappa}.$$

$\square$

Proposition 6. *Let $\widetilde{m}_1$ be a fuzzy mass function and m_2 a Bayesian mass function. Then the orthogonal sum $\widetilde{m}_1 \oplus m_2$ is Bayesian and it is given by*

$$(\widetilde{m}_1 \oplus m_2)(\{\theta\}) = \frac{\sum_{i=1}^{f} \widetilde{F}_i(\theta) \widetilde{m}_1(\widetilde{F}_i) m_2(\{\theta\})}{\sum_{\theta' \in \Theta} \sum_{i=1}^{f} \widetilde{F}_i(\theta') \widetilde{m}_1(\widetilde{F}_i) m_2(\{\theta'\})} = \frac{pl_{\widetilde{m}_1}(\theta) m_2(\{\theta\})}{\sum_{\theta' \in \Theta} pl_{\widetilde{m}_1}(\theta') m_2(\{\theta'\})}.$$

21

Proof. As the normalized product of each focal set $\widetilde{F}_i$ of $\widetilde{m}_1$ with each focal set $\{\theta\}$ of m_2 equals $\{\theta\}$, the orthogonal $\widetilde{m}_1 \oplus m_2$ is Bayesian. The expression of $(\widetilde{m}_1 \oplus m_2)(\{\theta\})$ follows directly from Proposition 5. $\qquad\square$

In DS theory, it is well known that Dempster's rule extends Bayesian conditioning, i.e., the orthogonal sum of a Bayesian belief function P and a logical belief function focussed on some subset A is a Bayesian belief function that is identical to the conditional probability measure $P(\cdot \mid A)$ [42]. In a similar way, the generalized product-intersection rule (48) extends conditioning of a probability measure by a fuzzy event, as stated in the following proposition.

Proposition 7. *Let m be a Bayesian mass function with corresponding probability measure P and $\widetilde{m}$ a logical fuzzy mass function with focal set $\widetilde{A} \in \mathcal{F}(\Theta)$. Then $m \oplus \widetilde{m}$ is Bayesian and the corresponding belief function $Bel_{m \oplus \widetilde{m}}$ is identical to the probability measure $P(\cdot \mid \widetilde{A})$ obtained by conditioning P by fuzzy event $\widetilde{A}$.*

Proof. From Proposition 6, $m \oplus \widetilde{m}$ is Bayesian and it is given by

$$(m \oplus \widetilde{m})(\{\theta\}) = \frac{m(\{\theta\})\widetilde{A}(\theta)}{\sum_{\theta' \in \Theta} m(\{\theta'\})\widetilde{A}(\theta')} = \frac{P(\{\theta\})\widetilde{A}(\theta)}{\sum_{\theta' \in \Theta} P(\{\theta'\})\widetilde{A}(\theta')}.$$

Consequently, for any $A \subseteq \Theta$,

$$Bel_{m \oplus \widetilde{m}}(A) = \sum_{\theta \in A}(m \oplus \widetilde{m})(\{\theta\}) = \frac{\sum_{\theta \in A} P(\{\theta\})\widetilde{A}(\theta)}{\sum_{\theta' \in \Theta} P(\{\theta'\})\widetilde{A}(\theta')} = P(A \mid \widetilde{A}).$$

$\qquad\square$

Finally, we can remark that the disjunctive rule (15) can also be extended to fuzzy mass functions. The extension is simpler in this case, because there is no normalization. For any t-conorm $\perp$, we can define a disjunctive operator $\mathbb{U}_\perp$ as

$$(\widetilde{m}_1 \,\mathbb{U}_\perp\, \widetilde{m}_2)(\widetilde{F}) := \sum_{\widetilde{G} \cup_\perp \widetilde{H} = \widetilde{F}} \widetilde{m}_1(\widetilde{G})\widetilde{m}_2(\widetilde{H}), \tag{50}$$

where $\cup_\perp$ denotes the union of fuzzy sets based on t-conorm $\perp$. The operator $\mathbb{U}_\perp$ is associative for any t-conorm. To be consistent with the conjunctive combination (48), it makes sense to choose the probabilistic sum $u \perp v = u + v - uv$, which is dual to the product t-norm. As already noted in Section 2.3, disjunctive combination may be more appropriate than normalized conjunctive rules (47)-(48) when the evidence is highly conflicting, in which case the normalizations in (45), (47) and (48) may result in numerical instability.

3.3. Belief and plausibility of fuzzy events

As plausibility and belief functions are, respectively, mixtures of possibility and necessity measures, definitions for the degree of belief and the degree of plausibility of a fuzzy event follow naturally from corresponding definitions in the possibilistic framework. As explained in Section 2.3, there are two definitions for the possibility and the necessity of fuzzy events, based on Sugeno and Choquet integrals. Consequently, a belief function Bel induced by a fuzzy mass function $\widetilde{m}$ can be extended to the lattice $(\mathcal{F}(\Theta), \wedge, \vee)$ in, at least, two ways, which are described below.

Extension based on Sugeno integrals. Based on the definitions of the possibility, necessity and guaranteed possibility of fuzzy events (32), (33) and (37), the belief, plausibility and commonality functions induced by a fuzzy mass function m can be extended to fuzzy events as:

$$Bel_{\widetilde{m}}^{(S)}(\widetilde{A}) := \sum_{i=1}^{f} m_i N_{\widetilde{F_i}}^{(S)}(\widetilde{A}) = \sum_{i=1}^{f} m_i \min_{\theta \in \Theta} (\widetilde{A} \vee \widetilde{F_i^c})(\theta) \tag{51a}$$

$$Pl_{\widetilde{m}}^{(S)}(\widetilde{A}) := \sum_{i=1}^{f} m_i \Pi_{\widetilde{F_i}}^{(S)}(\widetilde{A}) = \sum_{i=1}^{f} m_i \max_{\theta \in \Theta} (\widetilde{A} \wedge \widetilde{F_i})(\theta), \tag{51b}$$

and

$$Q_{\widetilde{m}}^{(S)}(\widetilde{A}) := \sum_{i=1}^{f} m_i \Delta_{\widetilde{F_i}}^{(S)}(\widetilde{A}) = \sum_{i=1}^{f} m_i \min_{\theta \in \Theta} (\widetilde{A}^c \vee \widetilde{F})(\theta). \tag{51c}$$

Definitions (51a) and (51b) were first introduced by Zadeh in [58]. As the necessity measures $N_{\widetilde{F_i}}^{(S)}$ are belief functions in the lattice $(\mathcal{F}(\Theta), \wedge, \vee)$, so is $Bel_{\widetilde{m}}^{(S)}$.

Extension based on Choquet integrals. Using now definitions (35) and (39) based on Choquet integrals, we get the following extensions of function $Bel_{\widetilde{m}}$, $Pl_{\widetilde{m}}$ and $Q_{\widetilde{m}}$ to fuzzy events:

$$Bel_{\widetilde{m}}^{(C)}(\widetilde{A}) := \int_0^1 Bel_{\widetilde{m}}(^\alpha \widetilde{A}) d\alpha = \sum_{i=1}^{f} m_i N_{\widetilde{F_i}}^{(C)}(\widetilde{A}) = \sum_{i=1}^{f} m_i Bel_{m_{\widetilde{F_i}}}(\widetilde{A}) \tag{52a}$$

$$Pl_{\widetilde{m}}^{(C)}(\widetilde{A}) := \int_0^1 Pl_{\widetilde{m}}(^\alpha \widetilde{A}) d\alpha = \sum_{i=1}^{f} m_i \Pi_{\widetilde{F_i}}^{(C)}(\widetilde{A}) = \sum_{i=1}^{f} m_i Pl_{m_{\widetilde{F_i}}}(\widetilde{A}), \tag{52b}$$

and

$$Q_{\widetilde{m}}^{(C)}(\widetilde{A}) := \int_0^1 Q_{\widetilde{m}}(^\alpha \widetilde{A}) d\alpha = \sum_{i=1}^{f} m_i \Delta_{\widetilde{F_i}}^{(C)}(\widetilde{A}) = \sum_{i=1}^{f} m_i Q_{m_{\widetilde{F_i}}}(\widetilde{A}), \tag{52c}$$

where, as before, $m_{\widetilde{F_i}}$ is the consonant mass functions such that $pl_{m_{\widetilde{F_i}}} = \pi_{\widetilde{F_i}}$. Definitions (52a) and (52b) were first proposed by Yen [53].

We can remark that definitions (51) and (52) coincide when either all focal sets are crisp, or the event $\widetilde{A}$ is crisp. From (36), definitions (52) also extend (44), i.e., we have

$$Bel_{\widetilde{m}}^{(C)}(\widetilde{A}) = \int_0^1 Bel_{\alpha\widetilde{m}}(\widetilde{A})d\alpha, \tag{53a}$$

$$Pl_{\widetilde{m}}^{(C)}(\widetilde{A}) = \int_0^1 Pl_{\alpha\widetilde{m}}(\widetilde{A})d\alpha, \tag{53b}$$

and

$$Q_{\widetilde{m}}^{(C)}(\widetilde{A}) = \int_0^1 Q_{\alpha\widetilde{m}}(\widetilde{A})d\alpha \tag{53c}$$

for all $\widetilde{A} \in \mathcal{F}(\Theta)$. Overall, both (51) and (52) seem to provide sensible definitions to quantify the degrees of belief and plausibility of fuzzy events based on uncertain and fuzzy evidence. Yen [53] argued for (52) by showing that the Choquet integrals are more sensitive to small changes in the fuzzy focal sets $\widetilde{F_i}$ than the Sugeno integrals. This argument could actually be reversed, as membership functions of fuzzy sets are usually difficult to determine precisely in practice, and the relative insensitivity of the degrees of belief and plausibility to small perturbations of the membership functions of the focal sets could rather be considered as an advantage. On the other hand, the fact that, for instance, $Bel_{\widetilde{m}}^{(C)}(\widetilde{A})$ can be determined by computing $Bel_{\beta\widetilde{m}}({}^{\alpha}\widetilde{A})$ for all α and β, and then by integrating over α and β, is a nice property of the Choquet integral. More research is needed to find decisive arguments in favor of any of these two approaches.

4. Application to statistical inference

In the following, we apply the general framework outlined in Section 3 to statistical inference. The consonant likelihood-based belief function, originally proposed by Shafer [42] is first recalled in Section 4.1, together with its axiomatic justification provided in [12]. In Section 4.2, we show that, by adding one axiom and enlarging the solution space, we can justify the representation of sample information by a fuzzy mass function with a single focal set, equal to the relative likelihood function. Binomial inference is used as an illustrative example in Section 4.3.

4.1. Likelihood-based belief function

Let us consider a statistical inference problem in which the observable data $X \in \mathcal{X}$ is randomly generated according to a probability mass or density function $f(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter whose value is only known to belong to a finite set Θ. After observing a realization x of X, we wish to represent the information about $\boldsymbol{\theta}$ by a belief function $Bel(\cdot; x)$ on Θ. Different approaches to this problem have been proposed by several authors, generalizing either Bayesian inference [6, 8, 9] or frequentist concepts such as confidence regions and p-values [2, 16, 36, 35].

A particularly simple and appealing solution, proposed by Shafer [42], is to consider the *consonant* belief function whose contour function is equal to the relative likelihood:

$$pl_0(\theta; x) := \frac{L(\theta; x)}{\max_{\theta' \in \Theta} L(\theta'; x)}, \tag{54a}$$

where $L(\cdot; x) : \theta \to f(x; \theta)$ is the likelihood function. The corresponding consonant plausibility and belief functions are, thus, defined as:

$$Pl_0(A; x) = \max_{\theta \in A} pl_0(\theta; x) \tag{54b}$$

and

$$Bel_0(A; x) = 1 - Pl_0(A^c; x) = 1 - \max_{\theta \notin A} pl_0(\theta; x), \tag{54c}$$

for all $A \subseteq \Theta$. This likelihood-based belief function has several interesting properties discussed in [12]. In particular, combining $Bel_0(\cdot; x)$ with a Bayesian prior probability mass function (PMF) on Θ yields the Bayesian posterior PMF.

The consonant belief function (54), introduced in [42] on intuitive grounds, can be justified from the Least Commitment Principle (LCP) [12]. This principle states that, when several belief functions satisfy some requirements for a reasonable representation of a given belief state, the least informative should be chosen [47]. Here, we consider the following requirement, which states that combining $Bel_\Theta(\cdot; x)$ with a prior PMF $\pi(\theta)$ by Dempster's rule yields the Bayesian posterior.

Requirement 1 (Compatibility with Bayesian inference). *Let $\pi(\theta)$ be a prior PMF. Then*

$$\pi \oplus Bel(\cdot; x) = P(\cdot \mid x), \tag{55}$$

where $P(\cdot \mid x)$ is the PMF on Θ defined by

$$P(\theta \mid x) = \frac{L(\theta; x)\pi(\theta)}{\sum_{\theta' \in \Theta} L(\theta'; x)\pi(\theta')}$$

for all $\theta \in \Theta$.

To apply the LCP, we need a way to compare the "information content" of belief functions. Here, we consider the *Q-ordering relation* [22] defined as follows: given two belief functions Bel_1 and Bel_2, with commonality functions Q_1 and Q_2, Bel_1 is Q-less committed than Bel_2 iff for all $A \subseteq \Theta$, $Q_1(A) \geq Q_2(A)$. The following proposition was proved in [12].

Proposition 8. *Belief function $Bel_0(\cdot; x)$ defined by (54) is the Q-least committed belief function verifying Requirement 1.*

Proof. From Proposition 2, Requirement 1 implies that

$$Q(\{\theta\}) = pl(\theta) = cL(\theta; x),$$

25

for all $\theta \in \Theta$ and for some constant c. The largest admissible value of c is

$$c_0 = [\max_{\theta \in \Theta} L(\theta; x)]^{-1}.$$

Hence, the Q-least committed belief function verifying Requirement 1 must be such that

$$Q(\{\theta\}) = \frac{L(\theta; x)}{\max_{\theta' \in \Theta} L(\theta'; x)} = pl_0(\theta; x).$$

Now, for any commonality function Q, we have

$$Q(A) \leq \min_{\theta \in A} Q(\{\theta\})$$

for all $A \subseteq \Theta$. Consequently, we must have $Q(A) \leq \min_{\theta \in A} pl_0(\theta; x)$ for all $A \subseteq \Theta$. The commonality function Q_0 corresponding to Bel_0 verifies $Q_0(A) = \min_{\theta \in A} pl_0(\theta; x)$. Consequently, $Q_0(A) \geq Q(A)$ for all A and all Q verifying Requirement 1. $\square$

Proposition 8 means that $Bel_0(\cdot; x)$ is, in some sense, the least informative belief function verifying Requirement 1. If we see the belief function $Bel_0(\cdot; x)$ as being induced by a consonant mass function $m_0(\cdot; x)$, there is, however, a problem with this representation: if x and y are the outcomes of independent random experiments with the same parameter space Θ, then $m_0(\cdot; x, y)$ is not equal to the orthogonal sum of $m_0(\cdot; x)$ and $m_0(\cdot; y)$, even though the two pieces of evidence x and y are independent. Indeed, it seems reasonable to impose the following requirement.

Requirement 2 (Combination of independent outcomes). *Let x and y be the outcomes of independent random experiments with the same parameter space Θ. Let $m(\cdot; x)$, $m(\cdot; y)$ and $m(\cdot; x, y)$ denote the mass functions on Θ induced, respectively, by the observation of x, y and (x, y). Then*

$$m(\cdot; x, y) = m(\cdot; x) \oplus m(\cdot; y).$$

The fact that mass function $m_0(\cdot; x)$ fails to meet Requirement 2 led Shafer to eventually reject it as a rational representation of statistical evidence [44]. It can be shown [49, 28] that the only mass function $m(\cdot; x)$ verifying Requirements 1 and 2, as well as the strong likelihood principle (i.e., depending only on the relative likelihood) and some additional regularity properties, is Bayesian and has the following expression:

$$m(\{\theta\}; x) = \frac{L(\theta; x)}{\sum_{\theta' \in \Theta} L(\theta'; x)}.$$

Hence, Requirement 2 seems to rule out belief function Bel_0, as well as any other non-additive representation of statistical evidence, an argument used by Halpern and Fagin to question the usefulness of belief functions [28]. However, this conclusion does not hold if we enlarge the solution space to include fuzzy mass functions, as will be shown in the next section.

4.2. Representation of statistical information as a fuzzy mass function

Belief function $Bel_0(\cdot; x)$ defined by (54) is induced by a single crisp consonant mass function $m_0(\cdot; x)$, but it is also induced by fuzzy mass functions. In particular, it is induced by the fuzzy mass function $\widetilde{m}_0(\cdot; x)$ such that $\widetilde{m}_0(\widetilde{L}_x; x) = 1$, where $\widetilde{L}_x$ is the fuzzy subset of Θ with membership function

$$\widetilde{L}_x(\theta) := \frac{L(\theta; x)}{\max_{\theta' \in \Theta} L(\theta'; x)} = pl_0(\theta; x)$$

for all $\theta \in \Theta$. Fuzzy set $\widetilde{L}_x$ can be seen as the *fuzzy set of likely values of $\boldsymbol{\theta}$ after observing* x. It is clear that $\widetilde{m}_0(\cdot; x)$ meets Requirement 2. Indeed, if x and y are independent observations,

$$\widetilde{L}_{x,y} = \widetilde{L}_x \odot \widetilde{L}_y$$

and

$$\widetilde{m}_0(\cdot; x, y) = \widetilde{m}_0(\cdot; x) \oplus \widetilde{m}_0(\cdot; y).$$

As a consequence of Proposition 6, $\widetilde{m}_0(\cdot; x)$ also meets Requirement 1. Fuzzy mass function $\widetilde{m}_0(\cdot; x)$ thus seems to be an adequate representation of statistical evidence: it verifies both requirements, and the associated belief function is the Q-least committed subject to Requirement 1.

We can remark that there are other fuzzy mass functions that induce $Bel_0(\cdot; x)$. They are characterized in the following proposition.

Proposition 9. *A fuzzy mass function $\widetilde{m}$ on $\Theta = \{\theta_1, \ldots, \theta_q\}$ with focal sets $\widetilde{F}_1, \ldots, \widetilde{F}_f$ and masses $m_1, \ldots, m_f$ verifies $Bel_{\widetilde{m}} = Bel_0$ if and only if*

$$\sum_{i=1}^{f} m_i \widetilde{F}_i = \widetilde{L}_x \tag{56}$$

and there exists a permutation σ of $\{1, \ldots, q\}$ such that, for all $i \in \{1, \ldots, f\}$,

$$\widetilde{F}_i(\theta_{\sigma(1)}) \leq \widetilde{F}_i(\theta_{\sigma(2)}) \leq \ldots \leq \widetilde{F}_i(\theta_{\sigma(q)}). \tag{57}$$

Proof. From (42), condition (56) is necessary and sufficient to ensure that $pl_{\widetilde{m}}(\theta) = Q_{\widetilde{m}}(\{\theta\}) = pl_0(\theta)$ for all $\theta \in \Theta$. Now, for any $A \subseteq \Theta$,

$$Q_{\widetilde{m}}(A) = \sum_{i=1}^{f} m_i \min_{\theta \in A} \widetilde{F}_i(\theta)$$

If (57) holds, then, for any $A \subseteq \Theta$, there exists θ_A in Θ such that, for any $i \in \{1, \ldots, f\}$, $\widetilde{F}_i(\theta_A) = \min_{\theta \in A} \widetilde{F}_i(\theta)$. We then have

$$Q_{\widetilde{m}}(A) = \sum_{i=1}^{f} m_i \widetilde{F}_i(\theta_A) = \min_{\theta \in A} \sum_{i=1}^{f} m_i \widetilde{F}_i(\theta) = \min_{\theta \in A} Q_{\widetilde{m}}(\{\theta\}) = Q_0(A).$$

Conversely, assume that $Q_{\widetilde{m}}(A) = Q_0(A) = \min_{\theta \in A} pl_0(\{\theta\})$ for all $A \subseteq \Theta$. Let σ be a permutation of $\{1, \ldots, q\}$ such that

$$pl_0(\theta_{\sigma(1)}) \le pl_0(\theta_{\sigma(2)}) \le \ldots \le pl_0(\theta_{\sigma(q)}).$$

For any $k \in \{1, \ldots, q-1\}$, we have

$$Q(\{\theta_{\sigma(k)}\}) = Q(\{\theta_{\sigma(k)}, \theta_{\sigma(k+1)}\}),$$

i.e.,

$$\sum_{i=1}^{f} m_i \widetilde{F}_i(\theta_{\sigma(k)}) = \sum_{i=1}^{f} m_i \min\left(\widetilde{F}_i(\theta_{\sigma(k)}), \widetilde{F}_i(\theta_{\sigma(k+1)})\right),$$

which implies that $\widetilde{F}_i(\theta_{\sigma(k)}) \le \widetilde{F}_i(\theta_{\sigma(k+1)})$ for all k and all i. $\qquad\square$

There are, thus, infinitely many fuzzy mass functions $\widetilde{m}$ that induce Bel_0. Let $\widetilde{m}(\cdot; x)$ and $\widetilde{m}(\cdot; y)$ be fuzzy mass functions corresponding to $Bel_0(\cdot; x)$ and $Bel_0(\cdot; y)$, assumed to have, respectively, f and f' focal sets. Then, $\widetilde{m}(\cdot; x) \oplus \widetilde{m}(\cdot; y)$ will generally have ff' focal sets and will be different from $\widetilde{m}(\cdot; x, y)$, except if $f = f' = 1$. Consequently, Requirement 2 justifies the choice of $\widetilde{m}_0$ as a fuzzy mass function representation of Bel_0, which is the Q-least committed belief function satisfying Requirement 1.

4.3. Application to binomial inference

Let us consider an urn with a known number N of balls and an unknown proportion $\boldsymbol{\theta}$ of black balls. Thus, $\boldsymbol{\theta} \in \Theta = \{0, 1/N, 2/N, \ldots, (N-1)/N, 1\}$. Assume that we have observed x black balls in n draws with replacement. The likelihood function is

$$L(\theta; x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

Let $\widehat{\theta}_x$ be the maximum likelihood estimate:

$$\widehat{\theta}_x = \arg\max_{\theta \in \Theta} L(\theta; x).$$

The fuzzy set $\widetilde{L}_x$ of likely values of $\boldsymbol{\theta}$ after observing x has membership function

$$\widetilde{L}_x(\theta) = \frac{L(\theta; x)}{L(\widehat{\theta}_x; x)} = \left(\frac{\theta}{\widehat{\theta}_x}\right)^x \left(\frac{1-\theta}{1-\widehat{\theta}_x}\right)^{n-x}.$$

The statistical evidence is represented by mass function $\widetilde{m}_0(\cdot; x)$ such that $\widetilde{m}_0(\widetilde{L}_x; x) = 1$. By construction, combining $\widetilde{m}_0(\cdot; x)$ by Dempster's rule with a Bayesian prior yields the Bayesian posterior. For instance, combining $\widetilde{m}_0(\cdot; x)$ with a uniform prior yields the posterior PMF

$$p(\theta \mid x) \propto \theta^x (1-\theta)^{n-x}.$$

If we now perform a second random experiment in which we observe y black balls out of q draws, the combined evidence about θ is represented by the fuzzy mass function

$$m_0(\cdot, x, y) = m_0(\cdot, x) \oplus m_0(\cdot, y)$$

with the single fuzzy focal set $\widetilde{L}_{x,y} = \widetilde{L}_x \odot \widetilde{L}_y$.

Table 1: Coverage probabilities of confidence regions ${}^{c_\alpha}\widetilde{L}_X$ for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$, with $\theta_0 = 0.3$.

	Coverage probability		
$1 - \alpha$	$N = 100, n = 50$	$N = n = 100$	$N = n = 1000$
0.99	0.9927	0.9922	0.9894
0.95	0.9570	0.9525	0.9511
0.90	0.8825	0.9025	0.8955

Frequentist properties. The previous analysis applies *after* the random experiment has been performed and the number x of black balls has been observed. *Before* the experiment has been performed, the number X of black balls is a random variable with binomial distribution, and we can consider the random set $(\Omega_X, 2^{\Omega_X}, P_{X|\theta_0}, \widetilde{\Gamma})$, where $\Omega_X = \{0, 1, \ldots, n\}$ is the sample space of X, $P_{X|\theta_0}$ is the probability distribution of X for the true value θ_0 of the parameter, and $\widetilde{\Gamma}(x) = \widetilde{L}_x$. The corresponding fuzzy mass function $\widetilde{m}$ is given by

$$\widetilde{m}(\widetilde{L}_x) = P_{X|\theta_0}(\{x\}) = \binom{n}{x} \theta_0^x (1 - \theta_0)^{n-x}$$

for all $x \in \Omega_X$. The plausibility $pl_{\widetilde{m}}(\theta_0)$ of θ_0 is a random variable that take values $\widetilde{L}_x(\theta_0)$, for $x \in \Omega$, with probabilities $P_{\theta_0}(\{x\})$. If N is large enough so that $\boldsymbol{\theta}$ can be treated as a continuous parameter, we have, by Wilks' theorem [50],

$$-2 \log pl_{\widetilde{m}}(\theta_0) \xrightarrow{d} \chi_1^2$$

and $n \to \infty$. Consequently,

$$P(-2 \log pl_{\widetilde{m}}(\theta_0) \leq \chi_{1;1-\alpha}^2) = P(pl_{\widetilde{m}}(\theta_0) \geq \exp(-0.5\chi_{1;1-\alpha}^2)) \to 1 - \alpha,$$

i.e., the c_α-cut of $\widetilde{L}_X$, with $c_\alpha = \exp(-0.5\chi_{1;1-\alpha}^2)$, is an asymptotic confidence region for $\boldsymbol{\theta}$ at level $1 - \alpha$. This property is illustrated in Figure 1 for $\theta_0 = 0.3$ and three cases: (1) $N = 100$, $n = 10$, (2) $N = n = 100$ and $N = n = 1000$. The true confidence levels of regions ${}^{c_\alpha}\widetilde{L}_X$ for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$ are shown in Table 1.

Prediction. Assume that we have picked x black balls out of n draws with replacement, and we now want to predict the number Y of black balls that will be obtained in another random experiment in which r balls will be drawn with replacement from the same urn. This prediction problem was addressed in [30, 31] in the belief function framework. The approach introduced in [30] is to write Y as a function $\varphi(\boldsymbol{\theta}, \boldsymbol{U})$ of parameter $\boldsymbol{\theta}$ and a pivotal random variable $\boldsymbol{U}$ with known probability distribution. A *predictive belief function* on Y is then obtained by combining the relation $Y = \varphi(\boldsymbol{\theta}, \boldsymbol{U})$ with the belief function on $\boldsymbol{\theta}$ derived from the likelihood function. Here, we can write

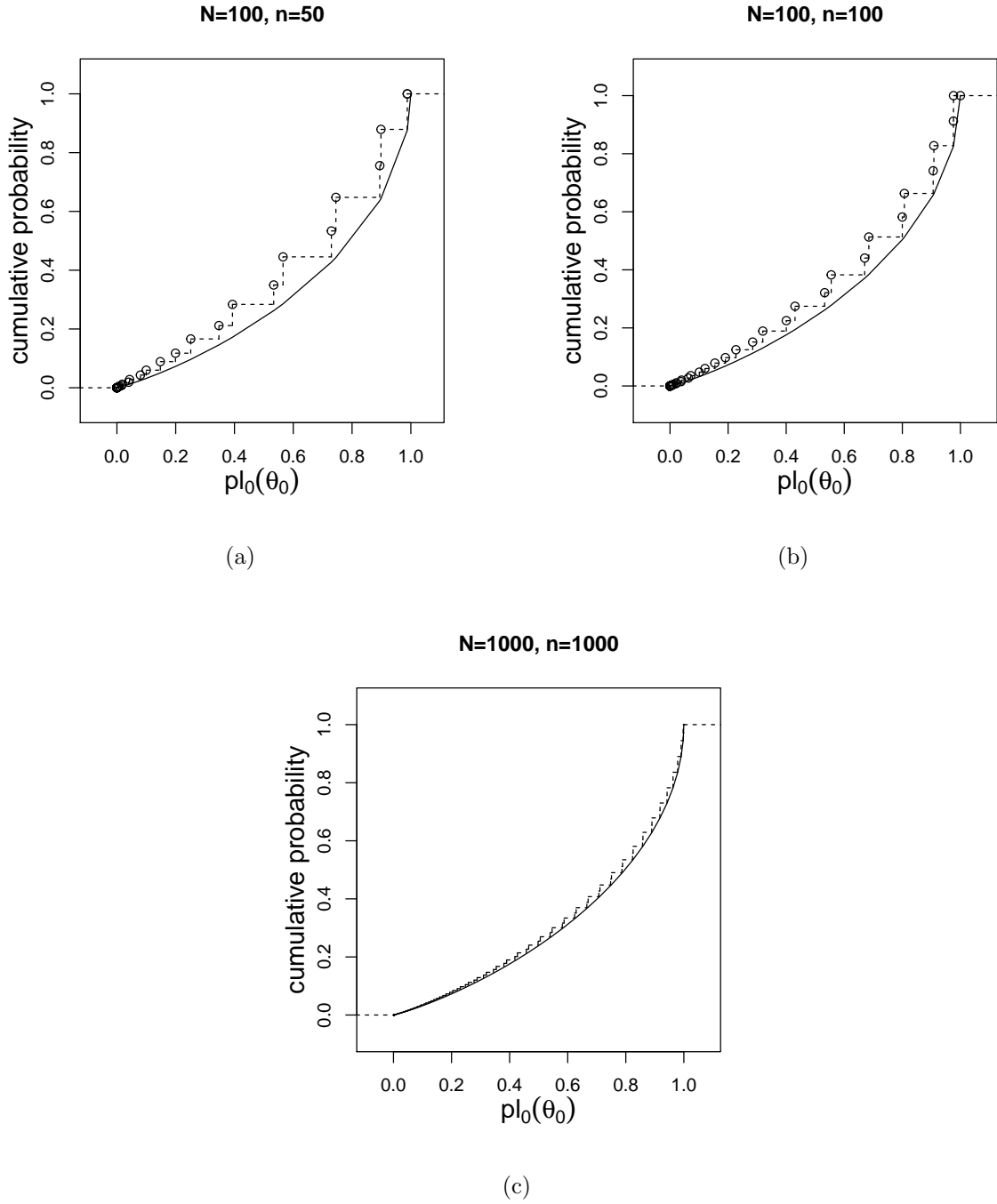$$Y = \sum_{i=1}^{r} I(U_i \leq \boldsymbol{\theta}), \tag{58}$$

Figure 1: True cdf of $pl_0(\theta_0; X) = \widetilde{L}_X(\theta)$ (broken lines) and asymptotic approximation (solid lines) for $N = 100$ and $n = 50$ (a), $N = n = 100$ (b) and $N = n = 1000$ (c).

where $\boldsymbol{U} = (U_1, \ldots, U_r)$ has a uniform distribution in $[0,1]^r$.

For a given realization $\boldsymbol{u} = (u_1, \ldots, u_r)$ of $\boldsymbol{U}$, the fuzzy constraint "$\boldsymbol{\theta}$ is $\widetilde{L}_x$" induces a fuzzy constraint "Y is $\widetilde{Y}_{x,\boldsymbol{u}}$", where $\widetilde{Y}_{x,\boldsymbol{u}}$ is a fuzzy subset of $\Omega_Y = \{0, \ldots, r\}$ whose membership function can be computed using the extension principle [56] as

$$\widetilde{Y}_{x,\boldsymbol{u}}(y) = \max_{\{\theta : y = \sum_{i=1}^r I(u_i \le \theta)\}} \widetilde{L}_x(\theta).$$

The mapping $\boldsymbol{u} \to \widetilde{Y}_{x,\boldsymbol{u}}$ and the probability distribution of $\boldsymbol{U}$ thus jointly define a random fuzzy set of Ω_Y. The corresponding fuzzy mass function $\widetilde{\mu}_x$ is defined by

$$\widetilde{\mu}_x(\widetilde{A}) = P(\{\boldsymbol{u} \in [0,1]^r : \widetilde{Y}_{x,\boldsymbol{u}} = \widetilde{A}\}),$$

for all $\widetilde{A} \in \mathcal{F}^*(\Omega_Y)$.

It can be shown that the predictive belief function generated by $\widetilde{\mu}_x$ also has an interesting frequentist property [16]. As before, let $c_\alpha = \exp(-0.5\chi^2_{1;1-\alpha})$, let ${}^{c_\alpha}\widetilde{\mu}_x$ be the (crisp) mass function whose focal sets are the c_α-cuts of the fuzzy focal sets of $\widetilde{\mu}_x$, and let ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}$ be the corresponding function. Then for any $\alpha \in (0,1)$, we have, asymptotically,

$$P_{X|\theta_0}\left(\left\{x \in \Omega_X : \forall A \subseteq \Omega_Y, \; {}^{c_\alpha}Bel_{\widetilde{\mu}_x}(A) \le P_{Y|\theta_0}(A)\right\}\right) \ge 1 - \alpha, \tag{59}$$

i.e., ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}$ dominates the true probability distribution $P_{Y|\theta_0}$ of Y for a proportion of the samples at least equal to $1 - \alpha$. Function ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}$ is, thus, a predictive belief function at confidence level $1 - \alpha$, according to the terminology introduced in [11].

As a numerical example, assume that we have observed $x = 28$ black balls out of $n = 100$ draws with replacement from an urn with $N = 100$ balls, and we wish to predict the number Y of black balls in $r = 4$ future draws from the same urn. The belief and plausibility of any event $A \subseteq \Omega_Y$ can be approximated by Monte Carlo simulation. More precisely, we generate K realizations $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K$ of $\boldsymbol{U}$, and we approximate $Pl_{\widetilde{\mu}_x}(A)$ and $Bel_{\widetilde{\mu}_x}(A)$ and, for any $A \subseteq \Omega_Y$ as, respectively

$$Pl_{\widetilde{\mu}_x}(A) \approx \frac{1}{K} \sum_{k=1}^K \max_{\theta \in A} \widetilde{Y}_{x,\boldsymbol{u}_k}(\theta)$$

and

$$Bel_{\widetilde{\mu}_x}(A) \approx 1 - \frac{1}{K} \sum_{k=1}^K \max_{\theta \notin A} \widetilde{Y}_{x,\boldsymbol{u}_k}(\theta).$$

In particular, the contour function is approximated as

$$pl_{\widetilde{\mu}_x}(\theta) \approx \frac{1}{K} \sum_{k=1}^K \widetilde{Y}_{x,\boldsymbol{u}_k}(\theta)$$

Figure 2a shows a plot of the contour function $pl_{\widetilde{\mu}_x}$ together with the plug-in PMF $p_{Y|\widehat{\theta}_x}$, with $\widehat{\theta}_x = x/n = 0.28$. Figure 2b displays the lower and upper cumulative distribution functions
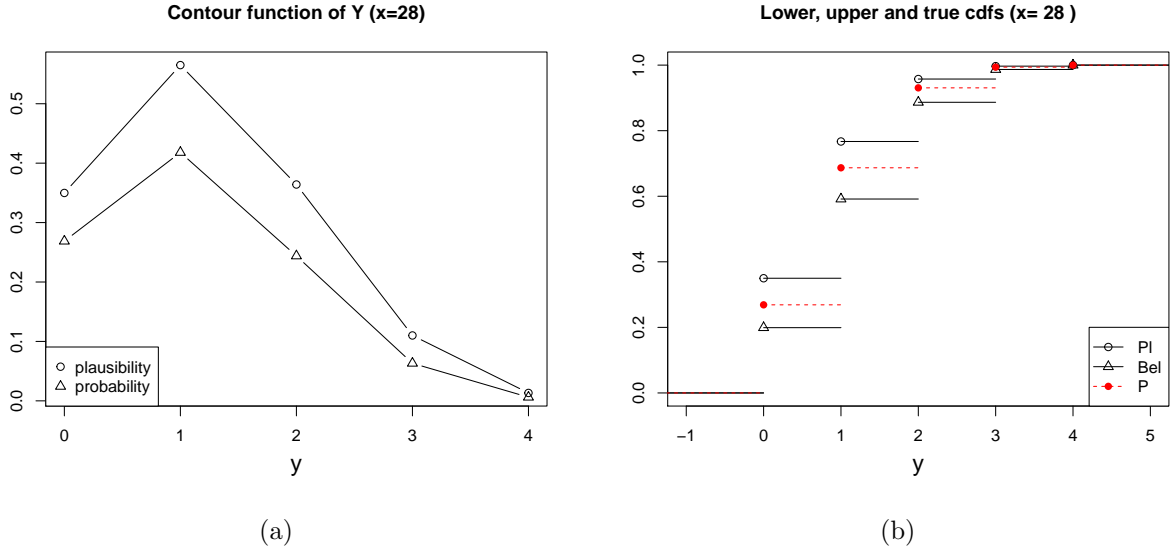
Figure 2: Views of the predictive belief function for $n = 100$, $x = 28$, $N = 100$ and $r = 4$. (a): Predictive contour function $pl_{\widetilde{\mu}_x}$ and true probability mass function $p_{Y|\theta_0}$ (with $\theta_0 = 0.3$); (b): Lower, upper and true CDFs.

Table 2: Estimated coverage probabilities of predictive belief functions ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}$ $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$, with $\theta_0 = 0.3$.

$1 - \alpha$	0.99	0.95	0.90
Coverage probability	0.9907	0.9496	0.8913

(CDFs) defined, respectively, as $F_*(y) = Bel_{\widetilde{\mu}_x}((-\infty, y])$ and $F^*(y) = Pl_{\widetilde{\mu}_x}((-\infty, y])$, together with the plug-in CDF $F_{Y|\widehat{\theta}_x}(y) = P_{Y|\widehat{\theta}}((-\infty, y])$. We can remark that, as $\widetilde{L}_x(\widehat{\theta}_x) = 1$, we have

$$Bel_{\widetilde{\mu}_x}(A) \leq P_{Y|\widehat{\theta}}(A) \leq Pl_{\widetilde{\mu}_x}(A)$$

for all $A \subseteq \Omega_Y$, which is verified in Figure 2b for $A = (-\infty, y]$ with $y \in \Omega_Y$.

To check the frequentist property (59) experimentally, we computed the coverage probability on the left-hand side of (59) for $\alpha \in \{0.01, 0.05, 0.1\}$. In this expression, ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}(A)$ can be approximated as

$$ {}^{c_\alpha}Bel_{\widetilde{\mu}_x}(A) \approx \frac{1}{K} \sum_{k=1}^{K} I({}^{c_\alpha}\widetilde{Y}_{x,\boldsymbol{u}_k} \subseteq A).$$

We set $K = 15,000$ in this simulation. The estimated probabilities are shown in Table 2. They are close to the nominal values $1 - \alpha$.

Figure 3 shows the belief degrees ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}(A)$ against the true probabilities $P_{Y|\theta_0}(A)$ for the $2^5 = 32$ subsets of Ω_Y and 9 values of x, with $\alpha = 0.05$. We can verify that we have, for the most probable values of x ($x \in \{22, 26, 30, 34, 38\}$), ${}^{c_\alpha}Bel_{\widetilde{\mu}_x}(A) \leq P_{Y|\theta_0}(A)$ for all A.
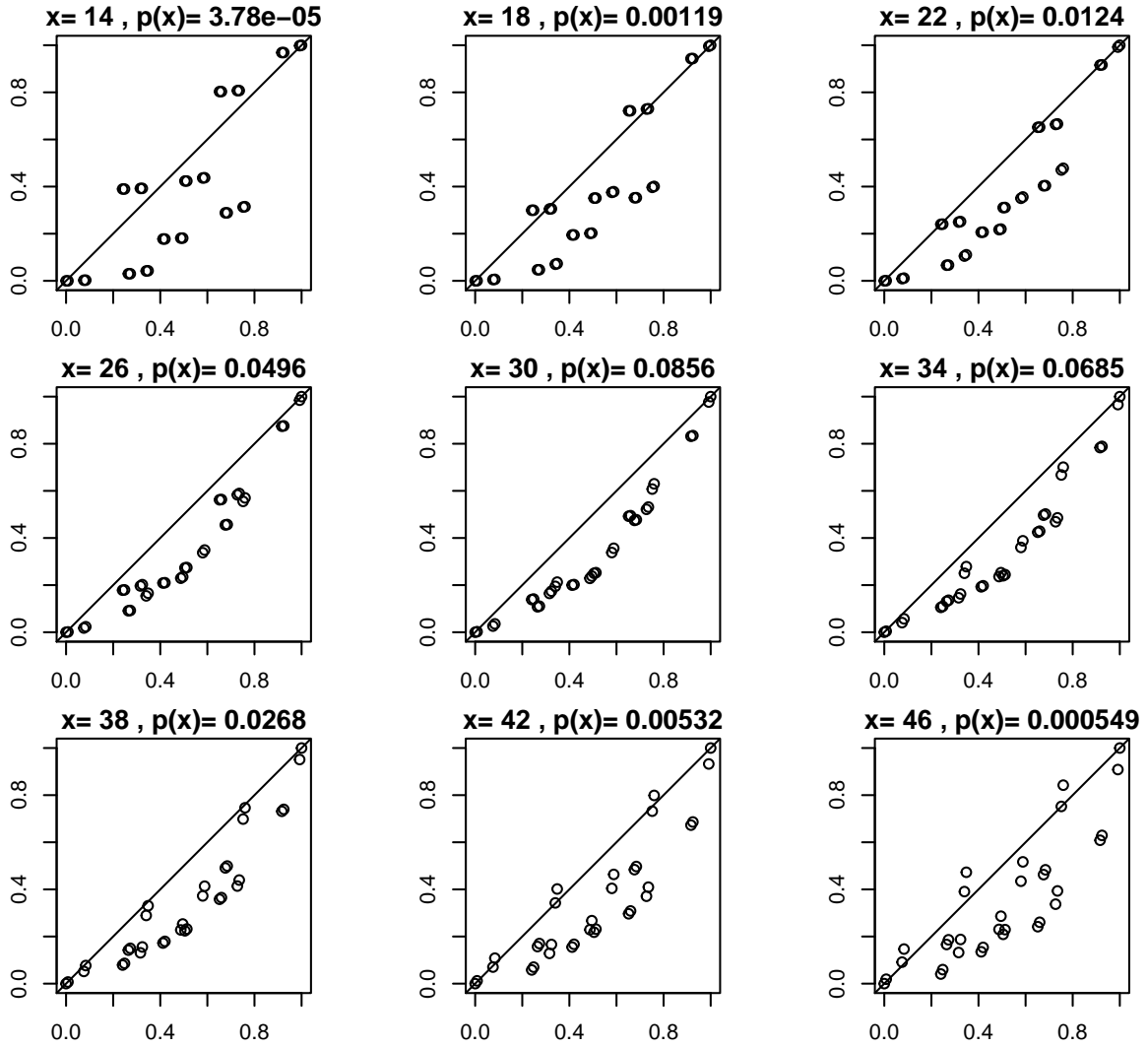
32

Figure 3: Degrees of belief $^{c_\alpha}Bel_{\widetilde{\mu}_x}(A)$ (vertical axes) vs. true probabilities $P_{Y|\theta_0}(A)$ (horizontal axes) or the $2^5 = 32$ subsets of $\Omega_Y = \{0, \dots, 4\}$ and 9 values of x, with $\alpha = 0.05$.

Probability of fuzzy events. The formalism exposed in Section 3.3 allows us to compute the belief and plausibility degrees of fuzzy events. For example, consider the fuzzy event: "Most balls will be black", where "most" corresponds to the following fuzzy subset of Ω_Y:

$$\widetilde{A} := \left\{ \frac{0}{0}, \frac{1}{0}, \frac{2}{0.5}, \frac{3}{0.75}, \frac{4}{1} \right\}.$$

Using Zadeh's definitions (51) for the belief and plausibility of fuzzy event based on the Sugeno integral, we get

$$Bel_{\widetilde{\mu}_x}^{(S)}(\widetilde{A}) \approx \frac{1}{K} \sum_{k=1}^{K} \min_{y \in \Omega_Y} (\widetilde{A} \vee \widetilde{Y}_{x,\boldsymbol{u}_k}^c)(y)$$

and

$$Pl_{\widetilde{\mu}_x}^{(S)}(\widetilde{A}) \approx \frac{1}{K} \sum_{k=1}^{K} \max_{y \in \Omega_Y} (\widetilde{A} \wedge \widetilde{Y}_{x,\boldsymbol{u}_k})(y).$$

For $x = 28$, we find $Bel_{\widetilde{\mu}_x}^{(S)}(\widetilde{A}) \approx 0.143$ and $Pl_{\widetilde{\mu}_x}^{(S)}(\widetilde{A}) \approx 0.253$. Using definitions (52) based on the Choquet integral, we have

$$Bel_{\widetilde{\mu}_x}^{(C)}(\widetilde{A}) = 0.5 Bel_{\widetilde{\mu}_x}^{(C)}(\{2,3,4\}) + 0.25 Bel_{\widetilde{\mu}_x}^{(C)}(\{3,4\}) + 0.25 Bel_{\widetilde{\mu}_x}^{(C)}(\{4\}) \approx 0.128$$

and

$$Pl_{\widetilde{\mu}_x}^{(C)}(\widetilde{A}) = 0.5 Pl_{\widetilde{\mu}_x}^{(C)}(\{2,3,4\}) + 0.25 Pl_{\widetilde{\mu}_x}^{(C)}(\{3,4\}) + 0.25 Pl_{\widetilde{\mu}_x}^{(C)}(\{4\}) \approx 0.236.$$

5. Conclusion

We have revisited Zadeh's notion of "evidence of the second kind" and outlined it importance as a generalization of both DS and possibility theories. According to this general view, possibility theory deals with belief functions generated by fuzzy sets, while DS theory deals with belief functions generated by random sets (or, equivalently, mass functions in the finite setting). Both theories fall under the umbrella of a more general formalism dealing with belief functions induced by random fuzzy sets, or fuzzy mass functions in the discrete case. This general formalism can be called "Epistemic random fuzzy set theory" (Figure 4). The adjective "epistemic" reminds us that, in this approach, a random fuzzy set is not seen as a model of a random mechanism for generating fuzzy data, but as a representation of evidence that can be both uncertain and fuzzy.

In this general approach, a belief function is only a partial representation of evidence, which is fully represented by the underlying random fuzzy set. In particular, two independent pieces of evidence must be combined at the random-fuzzy-set level. For instance, two possibility measures (or consonant belief functions) should be combined using the normalized product rule of possibility theory if they are generated by fuzzy sets, and by Dempster's rule is they are generated by random sets. These two combination mechanisms are special
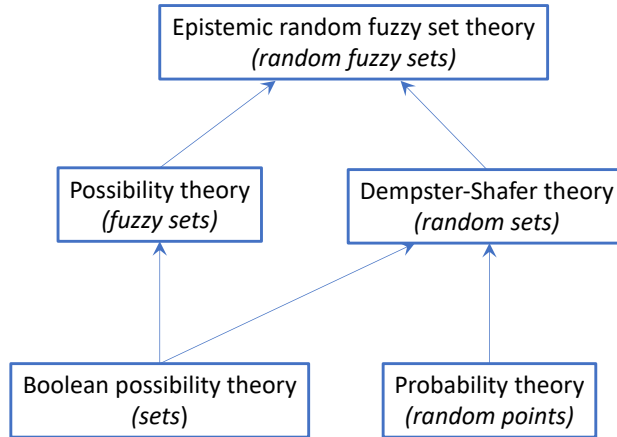
Figure 4: Overview of different theories of uncertainty, from the less general ones (bottom) to the most general (top).

cases of a generalized product-intersection rule, which allows us to combine independent pieces of evidence that are both uncertain and fuzzy.

As an application of this general framework, we have considered statistical inference. Shafer [42] first proposed to represent statistical evidence by the consonant belief function such that the plausibility of each single parameter value equals its relative likelihood. However, this representation was soon found to be problematic, as the belief function induced by the union of two independent samples is not equal to the orthogonal sum of the belief functions induced by each of the samples. This problem actually occurs only if we see the likelihood-based belief function as being generated by a crisp consonant mass function, i.e., if we see statistical evidence as uncertain. The problem disappears if we see statistical evidence as fuzzy, the relative likelihood function being interpreted as the membership function of the fuzzy set of likely values of the parameter. This perspective is consistent with the intuition of several authors who noticed the possibilistic nature of the concept of likelihood (see, e.g., [46]). Possibility theory alone, however, is not expressive enough to support the development of a theory of statistical inference as it does not allow us to represent, for instance, the combination of the likelihood with a Bayesian prior. The more general epistemic random fuzzy set theory allows us to reconcile the possibilistic interpretation of likelihood with Bayesian inference, and is rich enough to accommodate both "likelihoodist" and Bayesian views of statistical inference.

Throughout this paper, we have only considered finite sets to keep the level of mathematical exposition elementary. However, the whole framework developed in this paper can be extended to infinite spaces such as $\mathbb{R}^n$ without much difficulty, using existing results on random fuzzy sets in a more general setting (see, e.g., [5] and references therein). The practical application of such models to statistical inference or other uncertainty quantification problems will require the use of Monte Carlo simulation and numerical approximation techniques such as developed in [31, 48, 15]. Results in this direction will be reported in future publications.

References

[1] M. Aickin. Connecting Dempster-Shafer belief functions with likelihood-based inference. *Synthese*, 123:347–364, 2000.

[2] M. S. Balch. Mathematical foundations for a theory of confidence structures. *International Journal of Approximate Reasoning*, 53(7):1003 – 1019, 2012.

[3] C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, 45(1):82–105, 2007.

[4] I. Couso and L. Sánchez. Higher order models for fuzzy random variables. *Fuzzy Sets and Systems*, 159(3):237–258, 2008.

[5] I. Couso and L. Sánchez. Upper and lower probabilities induced by a fuzzy random variable. *Fuzzy Sets and Systems*, 165(1):1–23, 2011.

[6] A. P. Dempster. New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37:355–374, 1966.

[7] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[8] A. P. Dempster. A generalization of Bayesian inference (with discussion). *J. R. Statistical Society B*, 30:205–247, 1968.

[9] A. P. Dempster. The Dempster-Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48(2):365–377, 2008.

[10] T. Denœux. Modeling vague beliefs using fuzzy-valued belief structures. *Fuzzy Sets and Systems*, 116(2):167–199, 2000.

[11] T. Denœux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42(3):228–252, 2006.

[12] T. Denœux. Likelihood-based belief function: justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7):1535–1547, 2014.

[13] T. Denœux, D. Dubois, and H. Prade. Representations of uncertainty in artificial intelligence: Beyond probability and possibility. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research*, volume 1, chapter 4, pages 119–150. Springer Verlag, 2020.

[14] T. Denœux, D. Dubois, and H. Prade. Representations of uncertainty in artificial intelligence: Probability and possibility. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research*, volume 1, chapter 3, pages 69–117. Springer Verlag, 2020.

[15] T. Denœux and O. Kanjanatarakul. Multistep prediction using point-cloud approximation of continuous belief functions. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2019)*, New Orleans, USA, June 2019. IEEE.

[16] T. Denœux and S. Li. Frequency-calibrated belief functions: Review and new insights. *International Journal of Approximate Reasoning*, 92:232–254, 2018.

[17] T. Denœux and L. M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):47–62, 2001.

[18] D. Dubois, H. T. Nguyen, and H. Prade. Possibility theory, probability and fuzzy sets: Misunderstandings, bridges and gaps. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy sets*, pages 343–438. Kluwer Academic Publishers, Boston, 2000.

[19] D. Dubois, W. Ostasiewicz, and H. Prade. Fuzzy sets: History and basic notions. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy sets*, pages 21–124. Kluwer Academic Publishers, Boston, 2000.

[20] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In M. M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, New-York, 1982.

[21] D. Dubois and H. Prade. Evidence measures based on fuzzy information. *Automatica*, 21(5):547–562, 1985.

[22] D. Dubois and H. Prade. A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3):193–226, 1986.

[23] D. Dubois and H. Prade. *Possibility Theory: An approach to computerized processing of uncertainty.* Plenum Press, New-York, 1988.

[24] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In D. M. Gabbay and P. Smets, editors, *Handbook of Defeasible reasoning and uncertainty management systems*, volume 1, pages 169–226. Kluwer Academic Publishers, Dordrecht, 1998.

[25] D. Dubois, H. Prade, and R. Yager. Merging fuzzy information. In J. C. Bezdek, D. Dubois, and H. Prade, editors, *Fuzzy sets in approximate reasoning and information systems*, pages 335–401. Kluwer Academic Publishers, Boston, 1999.

[26] R. Féron. Ensembles aléatoires flous. *Compte-rendus de l'Académie des Sciences (Paris) – Série A*, 282:303–306, 1976.

[27] M. A. Gil, M. López-Díaz, and D. A. Ralescu. Overview on the development of fuzzy random variables. *Fuzzy Sets and Systems*, 157(19):2546–2557, 2006.

[28] J. Y. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54(3):275–317, 1992.

[29] M. Ishizuka and K. S. Fu. Inference procedures under uncertainty for the problem-reduction method. *Information Sciences*, 28:179–206, 1982.

[30] O. Kanjanatarakul, S. Sriboonchitta, and T. Denœux. Forecasting using belief functions: an application to marketing econometrics. *International Journal of Approximate Reasoning*, 55(5):1113–1128, 2014.

[31] O. Kanjanatarakul, S. Sriboonchitta, and T. Denœux. Statistical estimation and prediction using belief functions: principles and application to some econometric models. *International Journal of Approximate Reasoning*, 72:71–94, 2016.

[32] R. Kruse and K. D. Meyer. *Statistics with vague data.* Kluwer, Dordrecht, 1987.

[33] H. Kwakernaak. Fuzzy random variables – i. definitions and theorems. *Information Sciences*, 15(1):1–29, 1978.

[34] H. Kwakernaak. Fuzzy random variables – ii. algorithms and examples for the discrete case. *Information Sciences*, 17(3):253–278, 1979.

[35] R. Martin. False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, 113:39–73, 2019.

[36] R. Martin and C. Liu. *Inferential Models: Reasoning with Uncertainty.* CRC Press, Boca Raton, 2016.

[37] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.

[38] S. Petit-Renaud and T. Denœux. A fuzzy-neuro system for reconstruction of multi-sensor information. In W. Brauer, editor, *Proceedings of Fuzzy-Neuro Systems'98*, pages 322–329, Munich, March 1998. Infix.

[39] S. Petit-Renaud and T. Denœux. Regression analysis using fuzzy evidence theory. In *Proceedings of FUZZ-IEEE'99*, volume 3, pages 1229–1234, Seoul, August 1999.

[40] S. Petit-Renaud and T. Denœux. Nonparametric regression analysis of uncertain and imprecise data using belief functions. *International Journal of Approximate Reasoning*, 35(1):1–28, 2004.

[41] M. L. Puri and D. A. Ralescu. Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, 114(2):409–422, 1986.

[42] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[43] G. Shafer. Constructive probability. *Synthese*, 48(1):1–60, 1981.

[44] G. Shafer. Belief functions and parametric models (with discussion). *J. Roy. Statist. Soc. Ser. B*, 44:322–352, 1982.

[45] P. Smets. The degree of belief in a fuzzy event. *Information Sciences*, 25:1–19, 1981.

[46] P. Smets. Possibilistic inference from statistical data. In *Second World Conference on Mathematics at the service of Man*, pages 611–613, Universidad Politecnica de Las Palmas, 1982.

[47] P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[48] L. Sui, P. Feissel, and T. Denœux. Identification of elastic properties in the belief function framework. *International Journal of Approximate Reasoning*, 101:69–87, 2018.

[49] P. Walley. Belief function representations of statistical evidence. *The Annals of Statistics*, 15(4):1439–1465, 1987.

[50] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1):60–62, 1938.

[51] R. R. Yager. Generalized probabilities of fuzzy events from fuzzy belief structures. *Information Sciences*, 28:45–62, 1982.

[52] R. R. Yager. On the normalization of fuzzy belief structure. *International Journal of Approximate Reasoning*, 14:127–153, 1996.

[53] J. Yen. Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):559–569, 1990.

[54] L. A. Zadeh. Fuzzy sets. *Inform. Control*, 8:338–353, 1965.

[55] L. A. Zadeh. Probability measures of fuzzy events. *J. Math. Analysis and Appl.*, 10:421–427, 1968.

[56] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning –I. *Information Sciences*, 8:199–249, 1975.

[57] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[58] L. A. Zadeh. Fuzzy sets and information granularity. In M. M. Gupta, R. K. Ragade, and R. R. Yager, editors, *Advances in Fuzzy Sets Theory and Applications*, pages 3–18. North-Holland, Amsterdam, 1979.