International Journal of Approximate Reasoning 90 (2017) 333–340

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

The likelihood interpretation as the foundation of fuzzy set theory ☆

Marco E.G.V. Cattaneo

School of Mathematics and Physical Sciences, University of Hull, Kingston upon Hull, HU6 7RX, UK

ARTICLE INFO

Article history:
Received 9 January 2017
Received in revised form 16 June 2017
Accepted 14 August 2017
Available online 22 August 2017

Keywords:
Fuzzy sets
Likelihood function
Fuzzy data
Measurement error
Fuzzy inference
Combination rules

ABSTRACT

In order to use fuzzy sets in real-world applications, an interpretation for the values of membership functions is needed. The history of fuzzy set theory shows that the interpretation in terms of statistical likelihood is very natural, although the connection between likelihood and probability can be misleading. In this paper, the likelihood interpretation of fuzzy sets is reviewed: it makes fuzzy data and fuzzy inferences perfectly compatible with standard statistical analyses, and sheds some light on the central role played by extension principle and α-cuts in fuzzy set theory. Furthermore, the likelihood interpretation justifies some of the combination rules of fuzzy set theory, including the product and minimum rules for the conjunction of fuzzy sets, as well as the probabilistic-sum and bounded-sum rules for the disjunction of fuzzy sets.

1. Introduction

As far as works on fuzzy set theory remain in the realm of pure mathematics, a precise interpretation for the values of membership functions is not needed. However, as soon as examples of application are included, such an interpretation is necessary, otherwise not only the membership functions themselves will be arbitrary, but also all rules applied to them will be unjustified [1–4]. Unfortunately, most works involving application examples of fuzzy sets do not specify any clear interpretation of their membership values, and as a consequence it is not completely clear what those fuzzy sets exactly represent, or why some particular rules and not others have been employed [5–29].

The present paper, which is an extended version of [30], studies a possible interpretation of the values of membership functions: the one in terms of likelihood. This is probably the oldest interpretation of fuzzy sets, and despite some initial controversy, it is relatively common today [31]. The controversy stemmed from the close connection between likelihood and probability, which can generate the impression that this interpretation equates fuzzy set theory with probability theory. However, the concepts of probability and likelihood were clearly distinguished by Fisher [32]: likelihood is simpler, more intuitive, and better suited to information fusion [33–38].

In [32] Fisher also clearly stated that likelihood is defined only up to a multiplicative constant. The consequences of this fact for the likelihood interpretation of fuzzy sets have been often overlooked in the literature, and are analyzed in the present paper. Furthermore, this paper investigates which rules of fuzzy set theory are implied by the likelihood interpreta-

☆ This paper is part of the Virtual special issue on Soft methods in probability and statistics, Edited by Barbara Vantaggi, Maria Brigida Ferraro, Paolo Giordani.

E-mail address: m.cattaneo@hull.ac.uk.

http://dx.doi.org/10.1016/j.ijar.2017.08.006

tion. These rules include the extension principle for fuzzy sets, the product and minimum rules for the conjunction of fuzzy sets, as well as the probabilistic-sum and bounded-sum rules for the disjunction of fuzzy sets.

Hence, if in applications of fuzzy set theory the likelihood interpretation is adopted, then not only the values of membership functions have a clear meaning, but the use of the above rules is justified, since these rules are implied by the considered interpretation. In this sense, the likelihood interpretation can be seen as the foundation of the version of fuzzy set theory based on these rules.

The paper is organized as follows. In Section 2 the likelihood interpretation is presented, and some of its connections with other interpretations are briefly discussed. Section 3 studies the consequences of the likelihood interpretation of fuzzy data, including a natural concept of independence for fuzzy sets, and the justification of an expression for the likelihood function induced by fuzzy data that appeared often in the literature [12,23,29,39,40], but was not clearly justified. Furthermore, Section 3 explores the connections of the likelihood interpretation of fuzzy data with the assumption that the data are coarsened at random [41,42], as well as with errors-in-variables models or measurement error models [43]. The implications of the likelihood interpretation of fuzzy sets are analyzed further in Section 4, which discusses in particular the interpretation of α-cuts as confidence intervals, as well as the correspondence between extension principle and profile likelihood. Finally, in Section 5 the combination rules for fuzzy sets mentioned above are derived from the likelihood interpretation, while the last section concludes the paper.

2. The likelihood interpretation

A fuzzy set is described by its membership function $\mu : \mathcal{X} \to [0, 1]$, where $\mathcal{X}$ is a nonempty (crisp) set [44]. A standard example is the fuzzy set representing the meaning of the word "tall" in relation to a man, where the elements of $\mathcal{X}$ are the possible values of a man's height in cm [5–7]. We can expect for instance that $\mu(180) > \mu(160)$, because the attribute "tall" fits better to a 180 cm man than to a 160 cm one. However, the concept of a fuzzy set as described by a real-valued membership function μ can only be used to model the reality if we have an (operational) interpretation for the numerical values of μ, allowing us to (more or less precisely) quantify our ideas.

In fact, a clear interpretation of membership functions should be the starting point of a theory of fuzzy sets that describes the real world, and all rules of the theory should be a consequence of the interpretation [1–4]. This is for example the case with the theory of probability, whose rules are a consequence of each of its interpretations (at least on finite spaces) [45, 46]. As suggested by this example, it is not necessary that the interpretation of fuzzy sets is unique, but only the rules that are implied by the considered interpretation should be used in applications. Moreover, in the same way as subjective interpretations of probability are based on analogies with games of chance, or on hypothetical situations involving bets or preferences among actions, an interpretation of fuzzy sets based on analogies or hypothetical situations can also be sufficient.

One of the first aspects to consider when discussing the interpretation of fuzzy sets is if they are used in an epistemic or ontic sense [23,47]. Fuzzy sets have an ontic interpretation when they are themselves the object of inquiry, while they have an epistemic interpretation when their membership function $\mu : \mathcal{X} \to [0, 1]$ only gives information about the real object of inquiry, which is the value of $x \in \mathcal{X}$. In this paper, we will only consider epistemic fuzzy sets, and focus on their interpretation in terms of likelihood.

The likelihood interpretation of a fuzzy set consists in interpreting its membership function $\mu : \mathcal{X} \to [0, 1]$ as the likelihood function *lik* on $\mathcal{X}$ induced by the observation of an event D:

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) \tag{1}$$

for all $x \in \mathcal{X}$, where $P(D \mid x)$ was the probability of the event D (before its realization) given the value of $x \in \mathcal{X}$.

For example, "John is tall" is a piece of information that can be modeled by a fuzzy set with membership function $\mu : \mathcal{X} \to [0, 1]$ with $\mu(x) \propto P(D \mid x)$, where the elements of $\mathcal{X}$ are the possible values of John's height in cm, and $P(D \mid x)$ is the conditional probability of the event D of getting the information that "John is tall" given that John's height is x cm. Hence, the exact meaning of the interpretation of fuzzy sets in terms of likelihood depends on the interpretation given to probability values, but as noted above, the choice of this interpretation does not affect the rules of probability theory.

More specifically, we can imagine that for each possible value of $x \in \mathcal{X}$, a person chosen at random from a given population is asked to assign to John one out of a given list of possible attributes such as "tall" or "very tall", when the only information about John is that his height is x cm. In this case, D would be the event that the person selects the attribute "tall", and $P(D \mid x)$ could represent subjective probabilities (i.e. our degrees of belief in the occurrence of D given x) or objective probabilities (i.e. the proportions of persons in the population that would have selected the attribute "tall", depending on x) [45,46]. Of course, the whole situation could also be only hypothetical, and the exact meaning of the resulting membership function (1) would also depend on other aspects (besides the interpretation of probability), such as which population was considered, which possible attributes were listed, and how exactly the question was posed [48]. Anyway, when we have a membership function, we can always interpret it as the likelihood induced by the observation of an event in a completely specified situation (real or hypothetical): this is the likelihood interpretation, and the rules of fuzzy set theory implied by it do not depend on the details of the situation.

The likelihood interpretation is probably the oldest interpretation of fuzzy sets [31]: it has been more or less explicitly used directly after [49] and even before [50,51] the mathematical concept of fuzzy set was introduced by Zadeh [44], and has later been studied in detail by several authors [52–64]. However, most of them interpreted membership functions μ in terms of (conditional) probability values $\mu(x) = P(D \mid x)$, instead of likelihood values $\mu(x) = lik(x \mid D)$. Historically, the subtle distinction between probability and likelihood confused several great minds, before the likelihood of $x \in \mathcal{X}$ was clearly defined by Fisher as *proportional* to the conditional probability of the observed event D given x [32,65,66].

The proportionality constant in the definition of $lik(x \mid D)$ can depend on anything but the value of $x \in \mathcal{X}$. The reason for defining the likelihood function lik only up to a multiplicative constant is that otherwise lik would strongly depend on irrelevant information. For example, in the above situation with now two persons chosen at random from a population, assume that they assign to John the attributes "tall" and "very tall", respectively (without knowing the choice of the other person). Let D_1 be the event that the first person selects "tall" and the second one "very tall", while D_2 is the event that the first person selects "very tall" and the second one "tall". Because of symmetry, $P(D_1 \mid x) = P(D_2 \mid x)$, and therefore $P(D_1 \cup D_2 \mid x) = 2\, P(D_i \mid x)$, where $D_1 \cup D_2$ is the event that one of the two persons selects "tall" and the other one "very tall". Hence, if we interpreted membership functions in terms of (conditional) probability values, then the resulting fuzzy set would change completely if we had or did not have the (irrelevant) additional information about which person said "tall" and which one "very tall", while this does not happen with the likelihood interpretation (the only difference between the two interpretations is that likelihood is defined only up to a multiplicative constant).

Interpreting fuzzy sets in terms of likelihood thus implies that proportional membership functions have the same meaning. Uniqueness of representation is recovered by assuming, as we will do in the rest of the paper and is often done anyway, that all fuzzy sets are *normalized*. That is, their membership functions $\mu : \mathcal{X} \to [0, 1]$ satisfy $\sup_{x \in \mathcal{X}} \mu(x) = 1$, and are thus uniquely determined by $\mu(x) \propto P(D \mid x)$. Surprisingly, very few authors seem to have somehow considered this important aspect of the likelihood interpretation, and not in a very explicit way [3,53,58,59].

The likelihood interpretation of fuzzy sets is strictly related to two other interpretations often discussed in the literature: the ones based on random sets and on imprecise probability, respectively. The random set interpretation of a (not necessarily normalized) fuzzy set consists in interpreting its membership function $\mu : \mathcal{X} \to [0, 1]$ as the coverage function of a random subset S of $\mathcal{X}$: that is, $\mu(x) = P(x \in S)$ for all $x \in \mathcal{X}$ [54–56,59,67–69]. Except for the issue of normalization, the random set interpretation can be seen as a special case of the likelihood interpretation, because the observation of the event $x \in S$ induces on $\mathcal{X}$ the likelihood function proportional to μ (technically, the event $x \in S$ depends on x, but this is a minor detail that can be amended in various ways).

The imprecise probability interpretation of a (normalized) fuzzy set consists in interpreting the corresponding possibility measure on $\mathcal{X}$ (i.e. the one whose restriction to singletons corresponds to the membership function of the fuzzy set) as an upper probability measure [2,54,56,59,68,70,71], although then it is not clear why we should limit ourselves to describe our ideas using only possibility measures instead of much more general upper probability measures. Anyway, the connection with the likelihood interpretation comes from the fact that when an event D is observed, the upper probability measure on $\mathcal{X}$ is updated by means of the induced likelihood function μ on $\mathcal{X}$.

3. Fuzzy data

A basic advantage of the likelihood interpretation of fuzzy sets is that it allows to directly obtain statistical inferences from fuzzy data. The only condition on the statistical methods used is that the data enter them through the likelihood function only. In particular, all methods from the likelihood and Bayesian approaches to statistics can be straightforwardly generalized to the case of fuzzy data.

As discussed in Section 2, the membership function of a fuzzy set $\mu(x) \propto P(D \mid x)$ is interpreted as the likelihood function induced by the observation of an event D. Now, if we have a probability distribution on $x \in \mathcal{X}$, depending on an unknown parameter $\theta \in \Theta$, then the observation of the event D induces also a likelihood function lik on Θ:

$$lik(\theta \mid D) \propto P(D \mid \theta) = \int_{\mathcal{X}} P(D \mid x)\, dP(x \mid \theta) \propto \int_{\mathcal{X}} \mu(x)\, dP(x \mid \theta) \tag{2}$$

for all $\theta \in \Theta$, where $P(D \mid x)$ is assumed to be a measurable function of x that does not depend on θ.

Zadeh [39] defined the probability of the fuzzy event described by a membership function $\mu : \mathcal{X} \to [0, 1]$ as the right-hand side of (2), without justifying this choice through a clear interpretation of the values of μ. The likelihood interpretation provides only a partial justification: the right-hand side of (2) is proportional to the probability of the event D that induced the fuzzy information described by μ, where the proportionality constant can depend on anything but θ (or x).

Anyway, the unknown proportionality constant is not needed in order to obtain statistical inferences based on the likelihood function (2). In fact, the likelihood interpretation of fuzzy data can be seen as a generalization from crisp to fuzzy sets of the assumption that the data are coarsened at random, which itself is a generalization of the assumption that the data are missing at random [41,42]. More precisely, data are coarsened at random to a crisp set with indicator (or membership) function $\mu : \mathcal{X} \to [0, 1]$ when the likelihood function on $\mathcal{X}$ induced by their observation is proportional to μ, where the proportionality constant can depend on anything but θ (or x). The likelihood interpretation is a direct generalization of this to fuzzy data, which can be seen as meaning that the fact that we have observed some particular fuzzy data does not give us any information besides the one described by the membership functions. On the other hand, fuzzy data themselves can also be interpreted as describing the information we get from data that are crisply coarsened, but not at random.

In [39] Zadeh introduced also the concept of probabilistic independence for fuzzy events, again without a clear justification. The likelihood interpretation clarifies another concept of independence, which is extremely important in fuzzy set theory: the concept of independence among the pieces of information described by different fuzzy sets, which is usually implicitly or explicitly assumed [4,64]. The pieces of information described by the membership functions $\mu_1, \dots, \mu_n : \mathcal{X} \to [0,1]$ with $\mu_i(x) \propto P(D_i \mid x)$ can be interpreted as independent when the events $D_1, \dots, D_n$ that induced them were conditionally independent given x. In this case, the joint fuzzy information is described by the membership function $\mu : \mathcal{X} \to [0,1]$ with

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) = \prod_{i=1}^{n} P(D_i \mid x) \propto \prod_{i=1}^{n} \mu_i(x) \tag{3}$$

for all $x \in \mathcal{X}$, where $D = D_1 \cap \dots \cap D_n$.

In particular, if $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$, the components x_i of $x = (x_1, \dots, x_n)$ are probabilistically independent (for all θ), and each piece of fuzzy information $\mu_i(x_i) \propto P(D_i \mid x)$ is about a different component of x, then the assumption of their independence is very natural, and by combining (2) and (3) we obtain

$$lik(\theta \mid D) \propto \int_{\mathcal{X}} \prod_{i=1}^{n} \mu_i(x_i)\, dP(x \mid \theta) = \prod_{i=1}^{n} \int_{\mathcal{X}_i} \mu_i(x_i)\, dP(x_i \mid \theta) \tag{4}$$

for all $\theta \in \Theta$. This likelihood function has been considered by several authors [12,23,29,40], but was only justified on the basis of Zadeh's rather arbitrary definition of the probability of a fuzzy event [39]. By contrast, the likelihood interpretation of fuzzy sets provides a sound justification for the likelihood function (4) induced by fuzzy data.

The likelihood function (4) induced by fuzzy data with membership functions $\mu_i : \mathcal{X}_i \to [0,1]$ is often too complex to be handled analytically [12], but this is nowadays a typical situation in the likelihood and Bayesian approaches to statistics [35,72]. In particular, $x_1, \dots, x_n$ play the role of unobserved variables in (4), and therefore the EM algorithm can be used to maximize the likelihood [23]. Several examples of numerical calculations of maximum likelihood estimates based on fuzzy data are given for instance in [23,29].

When the data are fuzzy numbers, in the sense that $\mathcal{X}_i \subseteq \mathbb{R}$, the likelihood function (4) can also be interpreted as resulting from an errors-in-variables model or measurement error model [43]. In this case, the value ξ_i of a proxy x_i^* is assumed to be observed instead of the value of the variable x_i, where $\xi_i \in \mathbb{R}$ is an arbitrarily chosen constant, while the measurement error $\varepsilon_i = x_i^* - x_i$ is random with density $f_i \propto \mu_i(\xi_i - \cdot)$ and independent of everything else. In this model, each fuzzy number $\mu_i(x_i) \propto f_i(\xi_i - x_i) \propto lik(x_i \mid x_i^* = \xi_i)$ describes the information about the unknown value of x_i obtained from the observed value of its proxy x_i^*, and the likelihood function $lik(\cdot \mid x_1^* = \xi_1, \dots, x_n^* = \xi_n)$ on Θ induced by these observations is the one in (4). The description of fuzzy data in terms of measurement errors is particularly useful when the various components combine well mathematically, as in the following example.

Example 1. Assume that $x_1, \dots, x_n$ is a sample from a normal distribution with known variance σ^2 and unknown expectation $\theta \in \mathbb{R}$, but we have only fuzzy data with membership functions $\mu_i(x_i) = \exp\left(-{}^{(x_i - \xi_i)^2}/_{(2\,\sigma_i^2)}\right)$, where ξ_i, σ_i are known constants. Then the proxy variables $x_1^*, \dots, x_n^*$ are independent, and each x_i^* is normally distributed with expectation θ and variance $\sigma^2 + \sigma_i^2$. Hence, the likelihood function induced by the fuzzy data is given by

$$lik(\theta \mid x_1^* = \xi_1, \dots, x_n^* = \xi_n) \propto \exp\left(-\tfrac{(\theta - \hat{\theta})^2}{2\,\tau^2}\right) \tag{5}$$

for all $\theta \in \mathbb{R}$, where the maximum likelihood estimate $\hat{\theta}$ is a weighted average of the centers ξ_i of the fuzzy numbers, with weights ${}^{\tau^2}/_{(\sigma^2+\sigma_i^2)}$ depending on their precision ${}^1/_{\sigma_i^2}$, while ${}^1/_{\tau^2} = \sum_{i=1}^n {}^1/_{(\sigma^2+\sigma_i^2)}$ is the precision of $\hat{\theta}$ (which is normally distributed with expectation θ and variance τ^2).

Besides the maximum likelihood estimate $\hat{\theta}$, for each $\alpha \in (0,1)$ we obtain a likelihood-based confidence interval for θ:

$$\left\{\theta \in \mathbb{R} : lik(\theta) > \alpha\, lik(\hat{\theta})\right\} = \left(\hat{\theta} \pm \tau\,\sqrt{-2\,\ln\alpha}\right), \tag{6}$$

with exact level $F_{\chi_1^2}(-2\,\ln\alpha)$, where $F_{\chi_1^2}$ is the cumulative distribution function of the chi-squared distribution with 1 degree of freedom. Alternatively, we can combine the likelihood function (5) induced by the fuzzy data with a Bayesian prior for θ, and base our conclusions on the resulting posterior. In particular, if the prior is a normal distribution with expectation ϑ_0 and variance ς_0^2, then the posterior is a normal distribution with expectation ϑ_1 and variance ς_1^2, where ϑ_1 is a weighted average of ϑ_0 and $\hat{\theta}$, with weights proportional to their precision ${}^1/_{\varsigma_0^2}$ and ${}^1/_{\tau^2}$, respectively, which add up to the posterior precision ${}^1/_{\varsigma_1^2} = {}^1/_{\varsigma_0^2} + {}^1/_{\tau^2}$.

Furthermore, since the sample $x_1, \dots, x_n$ was only imprecisely observed, we could be interested in its conditional distribution given the fuzzy data [73]. For each possible value of θ, this distribution is as follows: $x_1, \dots, x_n$ are independent, and each x_i is normally distributed with expectation $\hat{\theta}_i$ and variance $\hat{\tau}_i^2$, where $\hat{\theta}_i$ is a weighted average of θ and ξ_i, with weights proportional to ${}^1/_{\sigma^2}$ and ${}^1/_{\sigma_i^2}$, respectively, which add up to the precision ${}^1/_{\hat{\tau}_i^2} = {}^1/_{\sigma^2} + {}^1/_{\sigma_i^2}$. In the likelihood approach to statistics, the value of θ is unknown, but the maximum likelihood estimate of the conditional distribution of the sample $x_1, \dots, x_n$ given the fuzzy data is simply the above distribution when we replace θ with $\hat{\theta}$. By contrast, in the Bayesian approach to statistics we have a posterior distribution for θ, and thus in general $x_1, \ldots, x_n$ are not independent, but only conditionally independent given θ.

4. Fuzzy inference

Besides allowing the direct use of fuzzy data in statistical methods, the likelihood interpretation of fuzzy sets also leads naturally to fuzzy statistical inference. In fact, the likelihood function on Θ induced by the (fuzzy or crisp) data can be interpreted as the membership function $\mu : \Theta \to [0, 1]$ of a fuzzy set describing the information obtained from the data about the unknown value of the parameter $\theta \in \Theta$.

In particular, the α-cuts of this fuzzy set,

$$\{\theta \in \Theta : \mu(\theta) > \alpha\} \tag{7}$$

with $\alpha \in (0, 1)$, correspond to the likelihood-based confidence intervals (or regions) for θ, such as the ones in (6). Both α-cuts and likelihood-based confidence intervals are usually defined using the non-strict inequality, but the choice of the strict inequality in (6) and (7) provides a better agreement with the concept of profile likelihood function [35,74], which is of central importance in the likelihood approach to statistics, and corresponds to the extension principle [5–7,75], which is equally central in fuzzy set theory.

More precisely, the extension principle states that the image under a function $g : \Theta \to \Gamma$ of the fuzzy set described by μ is the fuzzy set with membership function $\mu_g : \Gamma \to [0, 1]$ with

$$\mu_g(\gamma) = \sup\left\{\mu(\theta) : \theta \in \Theta,\ g(\theta) = \gamma\right\} \tag{8}$$

for all $\gamma \in \Gamma$ (where $\sup \varnothing = 0$). The α-cuts of the fuzzy set described by μ_g correspond to the images under g of the α-cuts (7) of the fuzzy set described by μ:

$$\left\{\gamma \in \Gamma : \mu_g(\gamma) > \alpha\right\} = \{g(\theta) : \theta \in \Theta,\ \mu(\theta) > \alpha\} \tag{9}$$

for all $\alpha \in (0, 1)$. This simple result would not be valid in general if the inequalities in (7) and (9) were not strict, because the supremum in (8) is not necessarily always a maximum. Anyway, in the likelihood approach to statistics we have exactly the same concepts under other names: the sets (9) are the likelihood-based confidence intervals (or regions) for $\gamma = g(\theta)$, while μ_g is (proportional to) the profile likelihood function on Γ obtained from μ through g [35,74]. These concepts are illustrated in an example about the addition of two fuzzy numbers at the end of the current section, and will then be employed in a similar way in the next section.

A correspondence between α-cuts and (general) confidence intervals has also been suggested as an alternative interpretation of some fuzzy sets [17,76]. However, this interpretation is afflicted by the fact that confidence intervals are rather arbitrary constructs, and in particular do not usually satisfy the extension principle, when they are not likelihood-based confidence intervals. The interpretation of fuzzy sets in terms of likelihood-based confidence intervals (i.e. the likelihood interpretation) has the advantage of uniqueness, invariance, and general applicability, although a simple expression for the confidence level based on the chi-squared distribution, as in Example 1, is valid (exactly or asymptotically) only under some regularity conditions [77].

Since each value of $\theta \in \Theta$ corresponds to a probability measure $P(\cdot\,|\,\theta)$, a fuzzy set with membership function $\mu : \Theta \to [0, 1]$ can also be interpreted as a fuzzy probability measure [36,78]. This likelihood-based model of fuzzy probability bears important similarities to the Bayesian model of probability, and can be used as a basis for statistical inference and decision making [36,38,78]. It is also closely related to other models of fuzzy probability discussed in the literature, with either the imprecise probability interpretation of fuzzy sets briefly mentioned at the end of Section 2 [79–82], or no clear interpretation [10,17,20,25,28].

Example 2. Let $\mu_1, \mu_2 : \mathbb{R} \to [0, 1]$ be the membership functions of two fuzzy numbers, interpreted as the likelihood functions induced by the observation of the events D_1, D_2, respectively: $\mu_i(x_i) \propto P(D_i\,|\,x_i)$, with $x_1, x_2 \in \mathbb{R}$. As discussed in Section 3, if we assume that the pieces of information described by μ_1 and μ_2 are independent, then the joint fuzzy information is described by the membership function $\mu(x_1, x_2) = \mu_1(x_1)\,\mu_2(x_2)$ on $\mathbb{R}^2$. In this case, the membership function of the sum of the two fuzzy numbers corresponds to the profile likelihood function μ_g on $\mathbb{R}$ obtained from μ through the function $g(x_1, x_2) = x_1 + x_2$ on $\mathbb{R}^2$:

$$\mu_g(x) = \sup_{y \in \mathbb{R}} \left(\mu_1(y)\,\mu_2(x - y)\right) \tag{10}$$

for all $x \in \mathbb{R}$. Hence, the product-sum (10) of fuzzy numbers [83–85] is justified by the likelihood interpretation when the independence of the fuzzy numbers is assumed.

The question of how to add two fuzzy numbers is slightly more complicated when their independence is not assumed, because in this case the conditional probability of $D_1 \cap D_2$ given x_1, x_2 cannot in general be obtained from the ones of D_1 and D_2. Moreover, since likelihood functions are defined only up to a multiplicative constant, we only know that $P(D_i\,|\,x_i) = c_i\,\mu_i(x_i)$ for all $x_i \in \mathbb{R}$, where $c_1, c_2 \in (0, 1]$ can be interpreted as additional parameters that are orthogonal to x_1, x_2 in the terminology of the likelihood approach to statistics [35]. However, all that we know about the conditional probability of $D_1 \cap D_2$ given x_1, x_2 (and c_1, c_2) are the Fréchet bounds:

$$(c_1\,\mu_1(x_1) + c_2\,\mu_2(x_2) - 1) \vee 0 \leq P(D_1 \cap D_2 \,|\, x_1, x_2) \leq (c_1\,\mu_1(x_1)) \wedge (c_2\,\mu_2(x_2)) \tag{11}$$

for all $x_1, x_2 \in \mathbb{R}$. In order to completely determine this probability, we can introduce a further additional (infinitely dimensional) parameter $\pi : \mathbb{R}^2 \to [0, 1]$ such that $P(D_1 \cap D_2 \,|\, x_1, x_2)$ is a weighted average of the lower and upper Fréchet bounds (11), with weights $1 - \pi(x_1, x_2)$ and $\pi(x_1, x_2)$, respectively, for all $x_1, x_2 \in \mathbb{R}$. Now, the conditional probability of $D_1 \cap D_2$ given the value of the parameters corresponds to the membership function $\mu(x_1, x_2, c_1, c_2, \pi)$ of the joint fuzzy information, while the membership function of the sum of the two fuzzy numbers corresponds to the profile likelihood function μ_g on $\mathbb{R}$ obtained from μ through the function $g(x_1, x_2, c_1, c_2, \pi) = x_1 + x_2$:

$$\mu_g(x) = \sup_{y \in \mathbb{R}} (\mu_1(y) \wedge \mu_2(x - y)) \tag{12}$$

for all $x \in \mathbb{R}$, since the maximum over c_1, c_2, π is always attained at $c_1 = c_2 = \pi(x_1, x_2) = 1$. Hence, the minimum-sum (12) of fuzzy numbers [5–7,83,84] is justified by the likelihood interpretation when the independence of the fuzzy numbers is not assumed.

5. Information fusion

The theory of fuzzy sets is also a theory of information fusion. In particular, (3) shows that the product rule for the conjunction of independent pieces of information is a consequence of the likelihood interpretation of fuzzy sets (1). Although this is the only combination rule straightforwardly implied by the definition of likelihood function, the rules for other logical connectives, with or without the independence assumption, can be obtained through the concept of profile likelihood (i.e. the extension principle), in the same way as the rules for the addition of two fuzzy numbers were obtained in Example 2.

More precisely, let $\mu_1, \mu_2 : \mathcal{X} \to [0, 1]$ be two membership functions describing two pieces of information, respectively, for instance "John is tall" and "John is very tall". The likelihood interpretation of fuzzy sets consists in interpreting each membership function μ_i as the likelihood function on $\mathcal{X}$ induced by the observation of an event D_i. Hence, the conjunction and disjunction of the two pieces of information are described by the membership functions $\mu_{1\cap 2}$ and $\mu_{1\cup 2}$ on $\mathcal{X}$ corresponding to the likelihood functions induced by the observation of $D_1 \cap D_2$ and $D_1 \cup D_2$, respectively.

The conditional probabilities of $D_1 \cap D_2$ and $D_1 \cup D_2$ given $x \in \mathcal{X}$ can be obtained from the ones of D_1 and D_2 only in particular situations, such as when their independence is assumed. However, even in this case, the conditional probabilities of D_1 and D_2 given x cannot be obtained from μ_1 and μ_2, because likelihood functions are defined only up to a multiplicative constant: we only know that $P(D_i \,|\, x) = c_i\,\mu_i(x)$ for all $x \in \mathcal{X}$, where $c_1, c_2 \in (0, 1]$. The standard technique for getting rid of this kind of uncertainty is the same in the likelihood approach to statistics (the technique of profile likelihood) and in fuzzy set theory (the extension principle), consisting in the elimination of the nuisance parameters by taking the maximum (or supremum) of the likelihood or membership functions over them.

If the pieces of information described by μ_1 and μ_2 are independent, in the sense discussed in Section 3 (i.e. D_1 and D_2 are conditionally independent given x), then their conjunction and disjunction correspond to the product and sum of fuzzy sets [39,44], respectively:

$$\mu_{1\cap 2}(x) \propto \mu_1(x)\,\mu_2(x), \tag{13}$$

$$\mu_{1\cup 2}(x) = \mu_1(x) + \mu_2(x) - \mu_1(x)\,\mu_2(x), \tag{14}$$

for all $x \in \mathcal{X}$. The product rule for the independent conjunction of two fuzzy sets (13) is a direct consequence of the definition of (conditional) independence, and (3) shows that the exact values of c_1 and c_2 play no role in it. By contrast, they play an important role in the probabilistic-sum rule for the independent disjunction of two fuzzy sets (14), which is obtained by choosing the values of c_1 and c_2 that maximize its right-hand side: $c_1 = c_2 = 1$ (independently of x).

The same values of c_1 and c_2 maximize (independently of x) the right-hand sides of the minimum and bounded-sum rules obtained, respectively, for the conjunction and disjunction of fuzzy sets whose independence is not assumed:

$$\mu_{1\cap 2}(x) \propto \mu_1(x) \wedge \mu_2(x), \tag{15}$$

$$\mu_{1\cup 2}(x) = (\mu_1(x) + \mu_2(x)) \wedge 1, \tag{16}$$

for all $x \in \mathcal{X}$. Moreover, since the independence of the pieces of information described by μ_1 and μ_2 is not assumed, the maximum possible values of the conditional probabilities of $D_1 \cap D_2$ and $D_1 \cup D_2$ given x are used in the right-hand sides of (15) and (16), respectively, again in accordance with the standard technique for eliminating nuisance parameters in the likelihood approach to statistics and in fuzzy set theory.

The rule for the complement of a fuzzy set can be obtained in the same way: the negation of the piece of information represented by μ_1 is described by the membership function $\mu_{\neg 1}$ on $\mathcal{X}$ corresponding to the likelihood function induced by the observation of D_1^c. For example, if μ_1 describes the piece of information "John is tall", then $\mu_{\neg 1}$ describes the lack of this piece of information: this corresponds to "John is not tall" only when we are sure to get one of these two pieces of information (e.g. when we get an answer to the yes–no question "Is John tall?"). Anyway, even in this case the resulting fuzzy set is vacuous:

$$\mu_{\neg 1}(x) = 1, \tag{17}$$

for all $x \in \mathcal{X}$. This rule is obtained by taking the supremum of its right-hand side over all possible values of $c_1 \in (0, 1]$, which in this case corresponds to taking the limit $c_1 \to 0$.

Except for the independent conjunction rule (13), the derivation of all above rules from the likelihood interpretation of fuzzy sets involved getting rid of uncertainty about nuisance parameters through likelihood maximization. As a consequence, the rules (14)–(17) must be applied with care, and cannot in particular be freely combined, because the maximization step should be carried out only once at the end of the process. For instance, if we simply apply two times the negation rule (17) to μ_1 we do not recover μ_1, while this is the case when we postpone the (then irrelevant) maximization step until the end of the calculation, obtaining the likelihood function μ_1 induced by the observation of $(D_1^c)^c = D_1$.

6. Conclusion

In order to use fuzzy sets in real-world applications, an (operational) interpretation for the values of membership functions is needed. The interpretation does not need to be unique, and can be based on analogies or hypothetical situations. However, in each real-world application of fuzzy set theory it should be clear which interpretation is used, and only the rules implied by this interpretation should be employed.

In this paper, the likelihood interpretation of fuzzy sets has been reviewed and some of its consequences analyzed. It consists in interpreting membership functions as likelihood functions, and as such it is an epistemic interpretation, and for normalized fuzzy sets only (since proportional likelihood functions are equivalent). The history of fuzzy set theory shows that the likelihood interpretation is very natural, but the connection between likelihood and probability can be misleading. However, likelihood and probability are complementary descriptions of uncertainty.

Not surprisingly, with the likelihood interpretation fuzzy data and fuzzy inferences are perfectly compatible with standard statistical analyses. In particular, the likelihood interpretation of fuzzy data justifies the use of expression (4) for the induced likelihood function, and establishes a fruitful connection with errors-in-variables models or measurement error models, as illustrated by Example 1. Furthermore, the link between this interpretation and the likelihood approach to statistics sheds some light on the central role played by extension principle and α-cuts in fuzzy set theory.

Finally, the likelihood interpretation justifies some of the combination rules of fuzzy set theory, including the product (13) and minimum (15) rules for the conjunction of fuzzy sets (with and without the assumption of their independence, respectively), as well as the probabilistic-sum (14) and bounded-sum (16) rules for the disjunction of fuzzy sets (again with and without the assumption of their independence, respectively).

References

[1] G. Shafer, Belief functions and possibility measures, in: J.C. Bezdek (Ed.), Analysis of Fuzzy Information, vol. I: Mathematics and Logic, CRC, 1987, pp. 51–84.
[2] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.
[3] D.V. Lindley, Comment to [58], J. Am. Stat. Assoc. 99 (2004) 877–879.
[4] J. Bradley, Fuzzy logic as a theory of vagueness: 15 conceptual questions, in: R. Seising (Ed.), Views on Fuzzy Sets and Systems from Different Perspectives, Springer, 2009, pp. 207–228.
[5] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—I, Inf. Sci. 8 (1975) 199–249.
[6] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—II, Inf. Sci. 8 (1975) 301–357.
[7] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—III, Inf. Sci. 9 (1975) 43–80.
[8] S.R. Watson, J.J. Weiss, M.L. Donnell, Fuzzy decision analysis, IEEE Trans. Syst. Man Cybern. 9 (1979) 1–9.
[9] R.R. Yager, Possibilistic decisionmaking, IEEE Trans. Syst. Man Cybern. 9 (1979) 388–392.
[10] L.A. Zadeh, Fuzzy probabilities, Inf. Process. Manag. 20 (1984) 363–372.
[11] M.L. Puri, D.A. Ralescu, Fuzzy random variables, J. Math. Anal. Appl. 114 (1986) 409–422.
[12] M.Á. Gil, M.R. Casals, An operative extension of the likelihood ratio test from fuzzy data, Stat. Pap. 29 (1988) 191–203.
[13] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic, Prentice Hall, 1995.
[14] B. Kim, R.R. Bishu, Evaluation of fuzzy linear regression models by comparing membership functions, Fuzzy Sets Syst. 100 (1998) 343–352.
[15] M. Inuiguchi, J. Ramík, Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem, Fuzzy Sets Syst. 111 (2000) 3–28.
[16] P. D'Urso, Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data, Comput. Stat. Data Anal. 42 (2003) 47–72.
[17] J.J. Buckley, Fuzzy Probability and Statistics, Springer, 2006.
[18] M.Á. Gil, M. López-Díaz, D.A. Ralescu, Overview on the development of fuzzy random variables, Fuzzy Sets Syst. 157 (2006) 2546–2557.
[19] H.T. Nguyen, B. Wu, Fundamentals of Statistics with Fuzzy Data, Springer, 2006.
[20] C. Baudrit, I. Couso, D. Dubois, Joint propagation of probability and possibility in risk analysis: towards a formal framework, Int. J. Approx. Reason. 45 (2007) 82–105.
[21] S.H. Choi, J.J. Buckley, Fuzzy regression using least absolute deviation estimators, Soft Comput. 12 (2008) 257–263.
[22] M.B. Ferraro, R. Coppi, G. González-Rodríguez, A. Colubi, A linear regression model for imprecise response, Int. J. Approx. Reason. 51 (2010) 759–770.
[23] T. Denœux, Maximum likelihood estimation from fuzzy data using the EM algorithm, Fuzzy Sets Syst. 183 (2011) 72–91.
[24] P. D'Urso, R. Massari, A. Santoro, Robust fuzzy regression analysis, Inf. Sci. 181 (2011) 4154–4174.
[25] R. Viertl, Statistical Methods for Fuzzy Data, Wiley, 2011.
[26] I. Georgescu, Possibility Theory and the Risk, Springer, 2012.

[27] D. Wang, P. Zhang, L. Chen, Fuzzy fault tree analysis for fire and explosion of crude oil tanks, J. Loss Prev. Process Ind. 26 (2013) 1390–1398.
[28] I. Couso, D. Dubois, L. Sánchez, Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables, Springer, 2014.
[29] H.-Y. Jung, W.-J. Lee, J.H. Yoon, S.H. Choi, Likelihood inference based on fuzzy data in regression model, in: SCIS & ISIS 2014, IEEE, 2014, pp. 1175–1179.
[30] M. Cattaneo, The likelihood interpretation of fuzzy data, in: M.B. Ferraro, P. Giordani, B. Vantaggi, M. Gagolewski, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Data Science, Springer, 2017, pp. 113–120.
[31] R. Seising, The Fuzzification of Systems, Springer, 2007.
[32] R.A. Fisher, On the "probable error" of a coefficient of correlation deduced from a small sample, Metron 1 (1921) 3–32.
[33] A.W.F. Edwards, Likelihood, 2nd edition, Johns Hopkins University Press, 1992.
[34] J.L. Ford, S. Ghose, The primitive uncertainty construct: possibility, potential surprise, probability and belief; some experimental evidence, Metroeconomica 49 (1998) 195–220.
[35] Y. Pawitan, In All Likelihood, Oxford University Press, 2001.
[36] M. Cattaneo, Fuzzy probabilities based on the likelihood function, in: D. Dubois, M.A. Lubiano, H. Prade, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Handling Variability and Imprecision, Springer, 2008, pp. 43–50.
[37] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, J. Timmer, Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, Bioinformatics 25 (2009) 1923–1929.
[38] M. Cattaneo, Likelihood decision functions, Electron. J. Stat. 7 (2013) 2924–2946.
[39] L.A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968) 421–427.
[40] X. Liu, S. Li, Cumulative distribution function estimation with fuzzy data: some estimators and further problems, in: R. Kruse, M.R. Berthold, C. Moewes, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Synergies of Soft Computing and Statistics for Intelligent Data Analysis, Springer, 2013, pp. 83–91.
[41] D.F. Heitjan, D.B. Rubin, Ignorability and coarse data, Ann. Stat. 19 (1991) 2244–2253.
[42] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, 2nd edition, Wiley, 2002.
[43] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement Error in Nonlinear Models, 2nd edition, Chapman & Hall/CRC, 2006.
[44] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[45] D. Gillies, Philosophical Theories of Probability, Routledge, 2000.
[46] D.P. Rowbottom, Probability, Wiley, 2015.
[47] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: making sense of fuzzy sets, Fuzzy Sets Syst. 192 (2012) 3–24.
[48] A. Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, Science 185 (1974) 1124–1131.
[49] V.I. Loginov, Probability treatment of Zadeh membership functions and their use in pattern recognition, Eng. Cybern. 4 (1966) 68–69.
[50] M. Black, Vagueness, Philos. Sci. 4 (1937) 427–455.
[51] K. Menger, Ensembles flous et fonctions aléatoires, C. R. Acad. Sci. 232 (1951) 2001–2003.
[52] E. Hisdal, Are grades of membership probabilities?, Fuzzy Sets Syst. 25 (1988) 325–348.
[53] Y.Y. Chen, Bernoulli trials: from a fuzzy measure point of view, J. Math. Anal. Appl. 175 (1993) 392–404.
[54] D. Dubois, S. Moral, H. Prade, A semantics for possibility theory based on likelihoods, J. Math. Anal. Appl. 205 (1997) 359–380.
[55] T. Bilgiç, I.B. Türkşen, Measurement of membership functions: theoretical and empirical work, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, Springer, 2000, pp. 195–230.
[56] D. Dubois, H.T. Nguyen, H. Prade, Possibility theory, probability and fuzzy sets, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, Springer, 2000, pp. 343–438.
[57] G. Coletti, R. Scozzafava, Conditional probability, fuzzy sets, and possibility: a unifying view, Fuzzy Sets Syst. 144 (2004) 227–249.
[58] N.D. Singpurwalla, J.M. Booker, Membership functions and probability measures of fuzzy sets, J. Am. Stat. Assoc. 99 (2004) 867–877.
[59] D. Dubois, Possibility theory and statistical reasoning, Comput. Stat. Data Anal. 51 (2006) 47–69.
[60] G. Coletti, B. Vantaggi, From comparative degrees of belief to conditional measures, in: S. Greco, R.A. Marques Pereira, M. Squillante, R.R. Yager, J. Kacprzyk (Eds.), Preferences and Decisions, Springer, 2010, pp. 69–84.
[61] G. Coletti, O. Gervasi, S. Tasso, B. Vantaggi, Generalized Bayesian inference in a fuzzy context: from theory to a virtual reality application, Comput. Stat. Data Anal. 56 (2012) 967–980.
[62] G. Coletti, B. Vantaggi, Inference with probabilistic and fuzzy information, in: R. Seising, E. Trillas, C. Moraga, S. Termini (Eds.), On Fuzziness, vol. 1, Springer, 2013, pp. 115–119.
[63] R. Scozzafava, The membership of a fuzzy set as coherent conditional probability, in: R. Seising, E. Trillas, C. Moraga, S. Termini (Eds.), On Fuzziness, vol. 2, Springer, 2013, pp. 631–635.
[64] B. Kovalerchuk, Probabilistic solution of Zadeh's test problems, in: A. Laurent, O. Strauss, B. Bouchon-Meunier, R.R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, vol. 2, Springer, 2014, pp. 536–545.
[65] A.W.F. Edwards, The history of likelihood, Int. Stat. Rev. 42 (1974) 9–15.
[66] A. Hald, On the history of maximum likelihood in relation to inverse probability and least squares, Stat. Sci. 14 (1999) 214–222.
[67] H.T. Nguyen, On modeling of linguistic information using random sets, Inf. Sci. 34 (1984) 265–274.
[68] G. De Cooman, Integration in possibility theory, in: M. Grabisch, T. Murofushi, M. Sugeno (Eds.), Fuzzy Measures and Integrals, Physica-Verlag, 2000, pp. 124–160.
[69] C.J. Geyer, G.D. Meeden, Fuzzy and randomized confidence intervals and P-values, Stat. Sci. 20 (2005) 358–366.
[70] P. Walley, Measures of uncertainty in expert systems, Artif. Intell. 83 (1996) 1–58.
[71] M. Cattaneo, On maxitive integration, Fuzzy Sets Syst. 304 (2016) 65–81.
[72] C.P. Robert, The Bayesian Choice, 2nd edition, Springer, 2001.
[73] O. Hryniewicz, Comparison of fuzzy and crisp random variables by Monte Carlo simulations, in: P. Grzegorzewski, M. Gagolewski, O. Hryniewicz, M.Á. Gil (Eds.), Strengthening Links Between Data Analysis and Soft Computing, Springer, 2015, pp. 13–20.
[74] M. Cattaneo, A. Wiencierz, Likelihood-based imprecise regression, Int. J. Approx. Reason. 53 (2012) 1137–1154.
[75] H.T. Nguyen, A note on the extension principle for fuzzy sets, J. Math. Anal. Appl. 64 (1978) 369–380.
[76] G. Mauris, Inferring a possibility distribution from very few measurements, in: D. Dubois, M.A. Lubiano, H. Prade, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Handling Variability and Imprecision, Springer, 2008, pp. 92–99.
[77] S.S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Stat. 9 (1938) 60–62.
[78] M. Cattaneo, A generalization of credal networks, in: T. Augustin, F.P.A. Coolen, S. Moral, M.C.M. Troffaes (Eds.), ISIPTA '09, SIPTA, 2009, pp. 79–88.
[79] R.F. Nau, Indeterminate probabilities on finite sets, Ann. Stat. 20 (1992) 1737–1767.
[80] P. Walley, Statistical inferences based on a second-order possibility distribution, Int. J. Gen. Syst. 26 (1997) 337–383.
[81] G. De Cooman, P. Walley, A possibilistic hierarchical model for behaviour under uncertainty, Theory Decis. 52 (2002) 327–374.
[82] G. De Cooman, A behavioural model for vague probability assessments, Fuzzy Sets Syst. 154 (2005) 305–358.
[83] D. Dubois, H. Prade, Additions of interactive fuzzy numbers, IEEE Trans. Autom. Control 26 (1981) 926–936.
[84] J.G. Dijkman, H. van Haeringen, S.J. de Lange, Fuzzy numbers, J. Math. Anal. Appl. 92 (1983) 301–341.
[85] R. Fullér, On product-sum of triangular fuzzy numbers, Fuzzy Sets Syst. 41 (1991) 83–87.