The Likelihood Interpretation
of Fuzzy Data

Marco E.G.V. Cattaneo

Abstract The interpretation of degrees of membership as statistical likelihood is probably the oldest interpretation of fuzzy sets. It allows in particular to easily incorporate fuzzy data and fuzzy inferences in statistical methods, and sheds some light on the central role played by extension principle and α-cuts in fuzzy set theory.

Keywords Fuzzy sets · Foundations · Likelihood function · Measurement error · Fuzzy inference

1 Introduction

Most works on fuzzy set theory do not give any precise interpretation for the values of membership functions. This is not a problem as far as the works remain in the realm of pure mathematics. However, as soon as examples of application are included an interpretation is needed, otherwise not only the membership functions are arbitrary, but also all rules applied to them are unjustified [3, 25, 32].

In this paper, the interpretation of the values of membership functions in terms of likelihood is reviewed. The concepts of probability and likelihood were clearly distinguished by Fisher [19]: likelihood is simpler, more intuitive, and better suited to information fusion [6, 8]. The likelihood interpretation of fuzzy sets is elucidated in Sect. 2, while Sect. 3 shows that it justifies an expression for the likelihood function induced by fuzzy data that appeared often in the literature [13, 20, 23, 26, 35], but without a clear justification. This likelihood function can also be interpreted as resulting from an errors-in-variables model or measurement error model [5], as will be illustrated by a simple example. Finally, Sect. 4 discusses the interpretation of α-cuts as confidence intervals, while the last section concludes the paper and outlines future work.

M. Cattaneo (✉)
Department of Mathematics, University of Hull, Kingston upon Hull, UK
e-mail: m.cattaneo@hull.ac.uk

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 [34]. 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 [36]. 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 interpretation for the numerical values of μ.

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 [3, 25, 32]. 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). As suggested by this example, it is not necessary that the interpretation is unique, but only the rules that are implied by the considered interpretation should be used in applications.

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 [13, 15]. 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:

$$\mu(x) = lik(x \,|\, D) \propto P(D \,|\, x)$$

for all $x \in \mathcal{X}$, where $P(D \,|\, x)$ 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 $\mu(x) \propto P(D \,|\, x)$, where the elements of $\mathcal{X}$ are the possible values of John's height in cm, and $P(D \,|\, x)$ is the probability of the event D of getting the information that "John is tall" when 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.

The likelihood interpretation is probably the oldest interpretation of fuzzy sets: it has been more or less explicitly used directly after [27] and even before [2, 29] the

mathematical concept of fuzzy set was introduced by Zadeh [34], and has later been studied in detail by several authors [1, 10–12, 14, 16, 17, 22, 24, 30, 31]. However, most of them interpreted membership functions μ in terms of 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 probability of the data D given x [18, 19, 21].

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, if two persons chosen at random from a population independently tell us that John is "tall" and "very tall", respectively, then the resulting fuzzy set should not change completely if we would or would not have the additional information that the first person said "tall" and the second one "very tall".

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 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 [14, 25, 31].

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 Sect. 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:

$$lik(\theta \mid D) \propto P(D \mid \theta) = \int_{\mathcal{X}} P(D \mid x) \, dP(x \mid \theta) \propto \int_{\mathcal{X}} \mu(x) \, dP(x \mid \theta) \quad (1)$$

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 [35] defined the probability of the fuzzy event described by a membership function $\mu : \mathcal{X} \to [0, 1]$ as the right-hand side of (1), 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 (1) 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).

In [35] 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 [3, 24]. The pieces of information described by the membership functions $\mu_1, \ldots, \mu_n : \mathcal{X} \to [0, 1]$ with $\mu_i(x) \propto P(D_i \mid x)$ can be interpreted as independent when the events $D_1, \ldots, D_n$ 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

$$\mu(x) = lik(x \mid D) \propto P(D \mid x) = \prod_{i=1}^{n} P(D_i \mid x) \propto \prod_{i=1}^{n} \mu_i(x) \tag{2}$$

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

In particular, if $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$, the components x_i of $x = (x_1, \ldots, x_n)$ 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 (1) and (2) we obtain

$$lik(\theta \mid D) \propto \int_{\mathcal{X}} \prod_{i=1}^{n} \mu_i(x_i) \, dP(x \mid \theta) = \prod_{i=1}^{n} \int_{\mathcal{X}_i} \mu_i(x_i) \, dP(x_i \mid \theta) \tag{3}$$

for all $\theta \in \Theta$. This likelihood function has been considered by several authors [13, 20, 23, 26], but only justified on the basis of Zadeh's rather arbitrary definition of the probability of a fuzzy event [35].

The likelihood function (3) induced by fuzzy data with membership functions $\mu_i : \mathcal{X}_i \to [0, 1]$ is often too complex to be handled analytically [20], but this is nowadays a typical situation in the likelihood and Bayesian approaches to statistics. In particular, $x_1, \ldots, x_n$ play the role of unobserved variables in (3), and therefore the EM algorithm can be used to maximize the likelihood [13]. Several examples of numerical calculations of maximum likelihood estimates based on fuzzy data are given for instance in [13, 23].

When the data are fuzzy numbers, in the sense that $\mathcal{X}_i \subseteq \mathbb{R}$, the likelihood function (3) can also be interpreted as resulting from an errors-in-variables model or measurement error model [5]. 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 $lik(\cdot \mid x_1^* = \xi_1, \ldots, x_n^* = \xi_n)$ on Θ induced by these observations is the one in (3). The description of fuzzy data in terms of measurement errors is

particularly useful when the various components combine well mathematically, as in the following simple example.

Example 1 Assume that $x_1, \ldots, x_n$ 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 $x_1^*, \ldots, x_n^*$ 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

$$lik(\theta \mid x_1^* = \xi_1, \ldots, x_n^* = \xi_n) \propto \exp\left(-\frac{(\theta - \hat{\theta})^2}{2\tau^2}\right) \tag{4}$$

for all $\theta \in \mathbb{R}$, where the maximum likelihood estimate $\hat{\theta}$ is the 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:

$$\left\{\theta \in \mathbb{R} : lik(\theta) > \alpha \, lik(\hat{\theta})\right\} = \left(\hat{\theta} \pm \tau \sqrt{-2 \ln \alpha}\right), \tag{5}$$

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 (4) induced by the fuzzy data with a Bayesian prior, and base our conclusions on the resulting posterior. In particular, if the prior is a normal distribution with expectation θ_0 and variance τ_0^2, then the posterior is a normal distribution with expectation θ_1 and variance τ_1^2, where θ_1 is the weighted average of θ_0 and $\hat{\theta}$, with weights proportional to their precision $1/\tau_0^2$ and $1/\tau^2$, respectively, while these add up to the posterior precision $1/\tau_1^2 = 1/\tau_0^2 + 1/\tau^2$.

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 (normalized) fuzzy set describing the information obtained from the data about the unknown value of the parameter $\theta \in \Theta$.

In particular, the likelihood-based confidence intervals (or regions) for θ, defined as in the left-hand side of (5) for all $\alpha \in (0, 1)$, correspond to the α-cuts of the fuzzy set with membership function μ. Both likelihood-based confidence intervals and α-cuts are usually defined using the non-strict inequality, but the choice of the strict

inequality in the definition provides a better agreement with the concept of profile likelihood function [9], which is of central importance in the likelihood approach to statistics, and corresponds to the extension principle [36], which is equally central in fuzzy set theory.

A correspondence between α-cuts and (general) confidence intervals has also been suggested as an alternative interpretation of some fuzzy sets [4, 28]. 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 [33].

Since each value of $\theta \in \Theta$ corresponds to a probability measure $P(\cdot \mid \theta)$, a fuzzy set with membership function $\mu : \Theta \to [0, 1]$ can also be interpreted as a fuzzy probability measure [6, 7]. 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 [6–8].

5 Conclusion

In this paper, the likelihood interpretation of fuzzy sets has been reviewed and some of its consequences analyzed. Not surprisingly, with this interpretation fuzzy data and fuzzy inferences can be easily incorporated in statistical methods. In particular, the likelihood interpretation of fuzzy data justifies the use of expression (3) 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.

The theory of fuzzy sets is also a theory of information fusion. However, only the product rule $\mu(x) \propto \prod_{i=1}^{n} \mu_i(x)$ for the conjunction of independent pieces of information is directly justified by the likelihood interpretation (2). 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). For example, the conjunction without independence assumption is then given by the minimum rule $\mu(x) \propto \bigwedge_{i=1}^{n} \mu_i(x)$, while negation always results in the vacuous membership function $\mu \equiv 1$. Such rules, which are a consequence of the likelihood interpretation of fuzzy sets, will be the topic of future work.

References

1. Bilgiç T, Türkşen IB (2000) Measurement of membership functions: theoretical and empirical work. In: Prade H, Dubois D (eds) Fundamentals of fuzzy sets. Springer, pp 195–230
2. Black M (1937) Vagueness. Philos Sci 4:427–455
3. Bradley J (2009) Fuzzy logic as a theory of vagueness: 15 conceptual questions. In: Seising R (ed) Views on fuzzy sets and systems from different perspectives. Springer, pp 207–228
4. Buckley JJ (2006) Fuzzy probability and statistics. Springer, New York
5. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models, 2nd edn. Chapman & Hall/CRC
6. Cattaneo M (2008) Fuzzy probabilities based on the likelihood function. In: Lubiano MA, Prade H, Gil MÁ, Grzegorzewski P, Hryniewicz O, Dubois D (eds) Soft methods for handling variability and imprecision. Springer, New York, pp 43–50
7. Cattaneo M (2009) A generalization of credal networks. In: Augustin T, Coolen FPA, Moral S, Troffaes MCM (eds) ISIPTA '09, SIPTA, pp 79–88
8. Cattaneo M (2013) Likelihood decision functions. Electron J Stat 7:2924–2946
9. Cattaneo M, Wiencierz A (2012) Likelihood-based imprecise regression. Int J Approx Reason 53:1137–1154
10. Coletti G, Scozzafava R (2004) Conditional probability, fuzzy sets, and possibility: a unifying view. Fuzzy Sets Syst 144:227–249
11. Coletti G, Vantaggi B (2010) From comparative degrees of belief to conditional measures. In: Squillante M, Yager RR, Kacprzyk J, Greco S, Marques Pereira RA (eds) Preferences and decisions. Springer, pp 69–84
12. Coletti G, Vantaggi B (2013) Inference with probabilistic and fuzzy information. In: Seising R, Trillas E, Moraga C, Termini S (eds) On fuzziness, vol 1. Springer, pp 115–119
13. Denœux T (2011) Maximum likelihood estimation from fuzzy data using the EM algorithm. Fuzzy Sets Syst 183:72–91
14. Dubois D (2006) Possibility theory and statistical reasoning. Comput Stat Data Anal 51:47–69
15. Dubois D, Prade H (2012) Gradualness, uncertainty and bipolarity: making sense of fuzzy sets. Fuzzy Sets Syst 192:3–24
16. Dubois D, Moral S, Prade H (1997) A semantics for possibility theory based on likelihoods. J Math Anal Appl 205:359–380
17. Dubois D, Nguyen HT, Prade H (2000) Possibility theory, probability and fuzzy sets. In: Prade H, Dubois D (eds) Fundamentals of fuzzy sets. Springer, pp 343–438
18. Edwards AWF (1974) The history of likelihood. Int Stat Rev 42:9–15
19. Fisher RA (1921) On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1:3–32
20. Gil MÁ, Casals MR (1988) An operative extension of the likelihood ratio test from fuzzy data. Stat Pap 29:191–203
21. Hald A (1999) On the history of maximum likelihood in relation to inverse probability and least squares. Stat Sci 14:214–222
22. Hisdal E (1988) Are grades of membership probabilities? Fuzzy Sets Syst 25:325–348
23. Jung HY, Lee WJ, Yoon JH, Choi SH (2014) Likelihood inference based on fuzzy data in regression model. In: SCIS & ISIS 2014, IEEE, pp 1175–1179
24. Kovalerchuk B (2014) Probabilistic solution of Zadeh's test problems. In: Laurent A, Strauss O, Bouchon-Meunier B, Yager RR (eds) Information processing and management of uncertainty in knowledge-based systems, vol 2. Springer, pp 536–545
25. Lindley DV (2004) Comment to [31]. J Am Stat Assoc 99:877–879
26. Liu X, Li S (2013) Cumulative distribution function estimation with fuzzy data: Some estimators and further problems. In: Berthold MR, Moewes C, Gil MÁ, Grzegorzewski P, Hryniewicz O, Kruse R (eds) Synergies of soft computing and statistics for intelligent data analysis. Springer, pp 83–91
27. Loginov VI (1966) Probability treatment of Zadeh membership functions and their use in pattern recognition. Eng Cybern 4:68–69

28. Mauris G (2008) Inferring a possibility distribution from very few measurements. In: Lubiano MA, Prade H, Gil MÁ, Grzegorzewski P, Hryniewicz O, Dubois D (eds) Soft methods for handling variability and imprecision. Springer, pp 92–99
29. Menger K (1951) Ensembles flous et fonctions aléatoires. C R Acad Sci 232:2001–2003
30. Scozzafava R (2013) The membership of a fuzzy set as coherent conditional probability. In: Seising R, Trillas E, Moraga C, Termini S (eds) On fuzziness, vol 2. Springer, pp 631–635
31. Singpurwalla ND, Booker JM (2004) Membership functions and probability measures of fuzzy sets. J Am Stat Assoc 99:867–877
32. Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman and Hall
33. Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 9:60–62
34. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
35. Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23:421–427
36. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Inf Sci 8:199–249, 8:301–357, 9:43–80