Statistical Reasoning with Set-Valued Information: Ontic vs. Epistemic Views

Didier Dubois[1]

Abstract Sets, hence fuzzy sets, may have a conjunctive or a disjunctive reading. In the conjunctive reading a (fuzzy) set represents an object of interest for which a (gradual rather than Boolean) composite description makes sense. In contrast disjunctive (fuzzy) sets refer to the use of sets as a representation of incomplete knowledge. They do not model objects or quantities, but partial information about an underlying object or a precise quantity. In this case the fuzzy set captures uncertainty, and its membership function is a possibility distribution. We call epistemic such fuzzy sets, since they represent states of incomplete knowledge. Distinguishing between ontic and epistemic fuzzy sets is important in information-processing tasks because there is a risk of misusing basic notions and tools, such as distance between fuzzy sets, variance of a fuzzy random variable, fuzzy regression, etc. We discuss several examples where the ontic and epistemic points of view yield different approaches to these concepts.

1 Introduction

Traditional views of engineering sciences aim at building a mathematical model of a real phenomenon, via a data set containing observations of the concerned phenomenon. This mathematical model is approximate in the sense that it is an imperfect copy of the reality it intends to account for, but it is often precise, namely it typically takes the form of a real-valued function that represents, for instance, the evolution of a quantity over time. Approaches vary according to the class of functions used. The oldest and most common class is the one of linear functions, but a lot of works dealing with non-linear models have appeared, for instance and prominently, using neural networks

[1] IRIT, CNRS and Université de Toulouse, France, dubois@irit.fr

C. Borgelt et al. (Eds.): Towards Advanced Data Analysis, STUDFUZZ 285, pp. 119–136.
springerlink.com

and fuzzy systems. These two techniques for constructing precise models have been merged to some extent due to the great similarity between the mathematical account of fuzzy rules and neurons, and their possible synergy due to the joint use of linguistic interpretability of fuzzy rules and learning capabilities of neural nets. While innovative with respect to older modeling techniques, these methods remain in the traditional school of producing a simplified and imperfect substitute of reality as observed via precise data.

Besides, there also exists a strong tradition of accounting for the non-deterministic aspect of many real phenomena subject to randomness in repeated experiments, including the noisy environment of measurement processes. Stochastic models enable to capture the general trends of populations of observed events through the use of probability distributions having a frequentist flavor. The probability measure attached to a quantity then reflects its variability through observed statistical data. Again in this approach, a stochastic model is a precise description of variability in physical phenomena.

More recently, with the emergence of Artificial Intelligence, but also in connection with more traditional human-centered research areas like Economics, Decision Analysis and Cognitive Psychology, the concern of reasoning about knowledge has emerged as a major paradigm [29]. While this topic has been mainly developed in the framework of classical or modal logic, due to the long philosophical tradition in this area, it has strongly affected the development of new uncertainty theories [20], and has led to a critique of probability theory as a unique framework for the representation of variability and belief. These developments question traditional views of modeling as representing reality independently of perception. They suggest a different approach that should also account for the cognitive limitations of our observations of reality. In other words, one might think of developing the epistemic approach to modeling. We call *ontic model* a precise representation of reality (however inaccurate it may be), and *epistemic model* a mathematical representation both of reality and the knowledge of reality, that explicitly accounts for the limited precision of our measurement capabilities. Typically, while the output of an ontic model is precise (but possibly wrong), an epistemic model delivers an imprecise output (hopefully consistent with the reality it accounts for). An epistemic model should of course be as precise as possible, given the available incomplete information, but it should also be as plausible as possible, avoiding unsupported arbitrary precision.

This paper discusses epistemic modeling in the context of set-based representations, and the mixing of variability and incomplete knowledge as present in recent works in fuzzy set-valued statistics.

2 Ontic vs. Epistemic Sets

A set S defined in extension, is often denoted by listing its elements, say, in the finite case $\{s_1, s_2, \ldots, s_n\}$. As pointed out in a recent paper [21] this representation, when it must be used in applications, is ambiguous. In some cases, a set represents a real complex lumped entity. It is then a conjunction of its elements. It is a precisely described entity made of subparts. For instance, a region in a digital image is a conjunction of adjacent pixels; a time interval spanned by an activity is the collection of instants where this activity takes place. In other cases, sets are mental constructions that represent incomplete information about an object or a quantity. In this case, a set is used as a disjunction of possible items, or of values of this underlying quantity, one of which is the right one. For instance I may only have a rough idea of the birth date of the president of some country, and provide an interval as containing this birth date. Such an interval is the disjunction of mutually exclusive elements. It is clear that the interval itself is subjective (it is my knowledge), has no intrinsic existence, even if it refers to a real fact. The use of sets representing imprecise values can be found for instance in interval analysis [39]. Another example is the set of models of a propositional knowledge base: only one of them reflects the real situation. Moreover this set is likely to change by acquiring more information.

Sets representing collections C of elements forming composite objects will be called *conjunctive*; sets E representing incomplete information states will be called *disjunctive*. A conjunctive set is the precise representation of an objective entity (philosophically it is a *de re* notion), while a disjunctive set only represents incomplete information (it is *de dicto*). We also shall speak of *ontic* sets, versus *epistemic* sets, in analogy with ontic vs. epistemic actions in cognitive robotics [30]. An ontic set C is the value of a set-valued variable X (and we can write $X = C$). An epistemic set E contains the ill-known actual value of a point-valued quantity x and we can write $x \in E$. A disjunctive set E represents the epistemic state of an agent, hence does not exist per se. In fact, when reasoning about an epistemic set it is better to handle a pair (x, E) made of a quantity and the available knowledge about it.

A value s inside a disjunctive set E is a possible candidate value for x, while elements outside E are considered impossible. Its characteristic function can be interpreted as a possibility distribution [56]. This distinction between conjunctive and disjunctive sets was made by Zadeh himself [57] distinguishing between set-valued attributes (like the set of sisters of some person) from ill-known single-valued attributes (like the unknown single sister of some person). This issue has been extensively discussed by Yager [53] and Dubois and Prade [17] for the study of incomplete conjunctive information (whose representation requires a disjunctive set of conjunctive sets).

An epistemic set (x, E) does not necessarily accounts for an ill-known deterministic value. An ill-known quantity may be deterministic or stochastic. For instance, the birth date of a specific individual is not a random variable even if it can be ill-known. On the other hand the daily rainfall in a specific place is a stochastic variable, since it can be modelled by a probability distribution. An epistemic set then captures in a rough way information about a population via observations. For instance, there is a sample space Ω, and x can be a random variable taking values on S, but the probability distribution on Ω is unknown. All that is known is that $x(\omega) \in E$, that is $P_x(E) = 1$ where P_x is the probability measure of x. In that case, E represents the family $\mathcal{P}_E$ of objective probability measures on Ω such that $P(\{\omega : x(\omega) \in E\}) = 1$, one of which being the proper representation of the random phenomenon. In this case, the object to which E refers is not a precise value of x, but a probability measure P_x describing the variability of x.

Note that in the probabilistic literature, an epistemic set is more often than not modelled by a probability distribution. In the early 19th century, Laplace proposed to use a uniform probability on E, based on the insufficient reason principle, according to which what is equipossible must be equiprobable. This is a default choice in $\mathcal{P}_E$ that coincides with the probability distribution having maximal entropy. However, this approach makes sense if x is a random variable. In case x is an ill-known deterministic value, Bayesians [35] propose to use a subjective probability P_x^b in place of set E. In that case, where the occurrence of x is not a matter of repetitions, $P_x^b(A)$ is the price of a lottery ticket chosen by an agent who agrees to earn \$1 if A turns out to be true, in an exchangeable bet scenario where the bookmaker exchanges roles with the buyer if the proposed price is found unfair. It forces the agent to propose prices $p^b(s)$ that sum exactly to 1 over E. Then $P_x^b(A)$ measures the degree of belief of the (non-repeatable) event $x \in E$, and this degree is agent-dependent.

However clever it may be, this view is debatable (see [20] for a summary of critiques). Especially, this representation is unstable: if P_x^b is uniform on E, then $P_{f(x)}^b$ may fail to be so if E is finite and the image $f(E)$ does not contain the same number of elements as E, or if E is an interval and f is not a linear transformation. Moreover, the use of unique probability distributions to represent belief is challenged by experimental results (like Ellsberg paradox [4]), which show that individuals do not make decisions based on expected utility in front of partial ignorance.

3 Random Sets vs. Ill-known Random Variables

As opposed to the case of an epistemic set representing an ill-known probability distribution, another situation is when the probability space (Ω, P) is available[1], but each realisation of the random variable is represented as a set. This case covers two situations:

[1] In this paper, we assume Ω is finite to avoid mathematical difficulties.

1. **Random conjunctive sets**: The random variable $X(\omega)$ is multi-valued and takes values on the power set of a set S. For instance, S is a set of spoken languages, and $X(\omega)$ is the set of languages spoken by an individual ω. Or $X(\omega)$ is an ill-known area of interest in some spatial domain, and ω is the outcome of an experiment to locate it. Then a probability distribution p_X is obtained over 2^S, such that $p_X(C) = P(X = C)$. It is known in the literature as a random set (Kendall [31], Matheron [38]). In our terminology this is a random conjunctive (or ontic) set.
2. **Ill-known random variables**: The random variable $x(\omega)$ takes values on S but its realisations are incompletely observed. It means that $\forall\omega \in \Omega$, all that is known is that $x(\omega) \in E = X(\omega)$ where X is a multiple-valued mapping $\Omega \to 2^S$ representing the disjunctive set of mappings (called selections) $\{x : \Omega \to S, \forall\omega, x(\omega) \in X(\omega)\} = \{x \in X\}$ for short. In other words the triple (Ω, P, X) is an epistemic model of the random variable x. This is the approach of Dempster [11] to imprecise probabilities. He uses this setting to account for a parametric probabilistic model P_θ on a set U of observables, where $\theta \in \Theta$ is an ill-known parameter but the probability distribution of a function $\phi(u, \theta) \in \Omega$ is known. Then $S = \Theta \times U$ and $X(\omega) = \{(\theta, u), \exists\theta, \phi(u, \theta) = \omega\}$. It is clear that for each ω, $X(\omega)$ is an epistemic set restricting, for each observation u the actual (deterministic) value θ.

Shafer [46] has proposed a non-statistical view of the epistemic random set setting, based on a subjective probability m over 2^S, formally identical to p_X. In this setting called the theory of evidence, $m(E)$ represents the subjective probability that all that is known of a deterministic quantity x is of the form $x \in E$. This is the case when an unreliable witness testifies that $x \in E$ and p is the degree of confidence of the receiver agent in the validity of the testimony. Then with probability $m(E) = p$, $x \in E$ is a reliable information. It means that the testimony is useless with probability $m(S) = 1 - p$ assigned to the empty information S. This view of probability was popular until the end of the 18th century (see [41] for details and a general model of unreliable witness). More generally the witness can be replaced by a measurement device or a message-passing entity with state space U, such that if the device is in state u then the available information is of the form $x \in E(u) \subseteq S$, and $p(u)$ is the subjective probability that the device is in state u [47].

The above discussions lay bare the difference between random conjunctive and disjunctive sets, even if they share the same mathematical model. In the first case one may compute precise probabilities that a set-valued variable X takes value in a family $\mathcal{A}$ of subsets:

$$P_X(\mathcal{A}) = \sum_{X(\omega) \in \mathcal{A}} p(\omega) = \sum_{C \in \mathcal{A}} p_X(C). \tag{1}$$

For instance, in the language example, and $S = \{\text{English, French, Spanish}\}$, one may compute the probability that someone speaks English by summing the proportions of people in Ω that respectively speak English only, English and French, English and Spanish, and the three languages.

In the second scenario, the random set $X(\omega)$ represents knowledge about a point-valued random variable $x(\omega)$. For instance, suppose S is an ordered height scale, $x(\omega)$ represents the height of individual ω and $X(\omega) = [a, b] \subseteq S$ is an imprecise measurement of $x(\omega)$. Here one can compute a probability range containing the probability $P_x(A) = \sum_{x(\omega)\in A} p(\omega)$ that the height of individuals in Ω lies in A, namely lower and upper probabilities proposed by Dempster [11]:

$$\underline{P}_X(A) = \sum_{X(\omega)\subseteq A} p(\omega) = \sum_{E\subseteq A} p_X(E); \tag{2}$$

$$\overline{P}_X(A) = \sum_{X(\omega)\cap A\neq\emptyset} p(\omega) = \sum_{E\cap A\neq\emptyset} p_X(E) \tag{3}$$

such that $\underline{P}_X(A) = 1 - \overline{P}_X(\bar{A})$, where $\bar{A}$ is the complement of A. Note that the set of probabilities $\mathcal{P}_X$ on S induced by this process is finite: since Ω and S are finite, the number of selections $x \in X$ is finite too. In particular, $\mathcal{P}_X$ is not convex. Its convex hull is $\tilde{\mathcal{P}}_X = \{P_S; \forall A \in S, P_S(A) \geq \underline{P}_X(A)\}$. It is well-known that probability measures in this convex set are of the form

$$P_S(A) = \sum_{E\subseteq S} p_X(E)P_E(A)$$

where P_E, a probability measure such that $P_E(E) = 1$, defines a sharing strategy of probability weight $p_X(E)$ among elements of E. As explained by Couso and Dubois [7], it corresponds to a scenario where when $\omega \in \Omega$ occurs, $x(\omega)$ is tainted with variability (due to the measurement device) that can be described by a conditional probability $P(\cdot|\omega)$ on S. Hence the probability $P_x(A)$ is now of the form:

$$P_x(A) = \sum_{\omega\in\Omega} P(A|\omega)P(\omega).$$

However, all we know is that $P(X(\omega)|\omega) = 1$ for some maximally specific epistemic subset $X(\omega)$. This is clearly a third (epistemic) view of the random set X. It is easy to see that the choice of $\mathcal{P}_X$ vs. its convex hull is immaterial in the computation of upper and lower probabilities, so that

$$\underline{P}_X(A) = \inf\{\sum_{\omega\in\Omega} P(A|\omega)P(\omega) : P(X(\omega)|\omega) = 1, \forall\omega \in \Omega\} \tag{4}$$

$$= \inf\{\sum_{E\subseteq S} p_X(E)P_E(A) : P_E(E) = 1\}. \tag{5}$$

where $P_E(A) = P(A|\omega)$ if $E = X(\omega)$.

In the evidence theory setting, Dempster upper and lower probabilities of an event are directly interpreted as degrees of belief $Bel(A) = \underline{P}_X(A)$ and plausibility $Pl(A) = \overline{P}_X(A)$, without reference to an ill-known probability on S (since the information is not frequentist here). This is the view of Smets [49]. The mathematical similarity between belief functions and random sets was quite early pointed out by Nguyen [40]. But they gave rise to quite distinct streams of literature that tend to ignore or misunderstand each other.

4 When the Meaning of the Model Affects Results

The reader may consider that the three above interpretations of random sets are just a philosophical issue, but do not impact on computations that can be carried out with this model. For instance the mean interval of a random interval has the same definition (interval arithmetics or Aumann integral) independently of the approach. However this is not true for other concepts. Two examples are given: conditioning and variance.

4.1 Conditioning Random Sets

Given a random set in the form of a probability distribution on the power set S, and an event $A \subset S$, the proper method for conditioning the random set on A depends on the adopted scenario.

Conditioning a conjunctive random set In this case the problem comes down to restricting the set-valued realisations $X(\omega)$ so as to account for the information that the set-valued outcome lies inside A. Then the obtained conditional random set is defined by means of the standard Bayes rule in the form of its weight distribution $p_X(\cdot|A)$ such that:

$$p_X(C|A) = \begin{cases} \frac{p_X(C)}{\sum_{B \subseteq A} p_X(B)} \text{ if } C \subseteq A; \\ 0 \text{ otherwise.} \end{cases} \tag{6}$$

Conditioning an ill-known random variable Suppose the epistemic random set $X(\omega)$ relies on a population Ω, and is represented by the convex set of probabilities $\tilde{\mathcal{P}}_X$ on S, one of which is the proper frequentist distribution of the underlying random variable x. Suppose we study a case for which all we know is that $x \in A$, and the problem is to predict the value of x. Each probability $p_X(E)$ should be altered in order to restrict to the subset $\Omega_A = \{\omega : x(\omega) \in A\}$ of population Ω. However, because $x(\omega)$ is only known to lie in $X(\omega)$, the set Ω_A is itself ill-known. There are three situations:

1. Either $A \cap E = \emptyset$: then $\Omega_A \cap \{\omega : X(\omega) = E\} = \emptyset$ and we can drop $p_X(E)$.
2. Or $E \subseteq A$ and then $\{\omega : X(\omega) = E\} \subseteq \Omega_A$ and $p_X(E)$ should remain assigned to E;
3. Or E overlaps both A and its complement: then let $\alpha_A(E)$ be the proportion of the population for which all we know is $x(\omega) \in E$ and that lies inside Ω_A. The weight $\alpha_A(E)p_X(E)$ should be assigned to $E \cap A$.

One may then define the conditional probability distribution over 2^S as follows:

$$p_X^{\alpha_A}(B|A) = \begin{cases} \frac{\sum_{B=E\cap A} \alpha_A(E)p_X(E)}{\sum_{E\cap A\neq\emptyset} \alpha_A(E)p_X(E)} & \text{if } B \subseteq A; \\ 0 \text{ otherwise.} \end{cases} \tag{7}$$

This mass assignment leads to computing lower and upper probabilities $\underline{P}^{\alpha_A}(\cdot|A)$ and $\overline{P}^{\alpha_A}(\cdot|A)$ when the vector of weights α_A is fixed. But this proportion $\alpha_A(E)$ is unknown in the third situation, while it is respectively 0 and 1 in the previous ones. Varying this unknown vector leads to upper and lower conditional probabilities as follows:

$$\overline{P}_X(B|A) = \sup_{\alpha_A} \overline{P}^{\alpha_A}(B|A); \qquad \underline{P}_X(B|A) = \inf_{\alpha_A} \underline{P}^{\alpha_A}(B|A). \tag{8}$$

and likewise for the lower conditional probability. In fact, it has been proved that these bounds can be obtained by applying Bayesian conditioning to all probabilities in $\tilde{\mathcal{P}}_X$ with $P_x(A) > 0$ and that they take an attractive closed form [9, 23]:

$$\overline{P}_X(B|A) = \sup\{P_x(B|A) : P_x \in \tilde{\mathcal{P}}_X\} = \frac{\overline{P}_X(B\cap A)}{\overline{P}_X(B\cap A) + \underline{P}_X(\bar{B}\cap A)}, \tag{9}$$

$$\underline{P}_X(B|A) = \inf\{P_x(B|A) : P_x \in \tilde{\mathcal{P}}_X\} = \frac{\underline{P}_X(B\cap A)}{\underline{P}_X(B\cap A) + \overline{P}_X(\bar{B}\cap A)}, \tag{10}$$

where $\underline{P}_X(B|A) = 1 - \overline{P}_X(\bar{B}|A)$ and $\bar{B}$ is the complement of B.

Conditioning a belief function In this case, there is no longer any population, and the probability distribution $m = p_X$ on 2^S represents subjective knowledge about a deterministic value x. Conditioning on A means that we come to hear that the actual value of x lies in A for sure. Then we perform an information fusion process (a special case of Dempster rule of combination [11]). It yields yet another type of conditioning, called Dempster conditioning, that systematically transfers masses $m(E)$ to $E \cap A$ when not empty, eliminates $m(E)$ otherwise, then normalises the conditional mass function, dividing by $\sum_{E\cap A\neq\emptyset} m(E) = Pl(A)$. It leads to the conditioning rule

$$Pl(B|A) = \frac{Pl(A\cap B)}{Pl(A)} = \frac{\overline{P}_X(A\cap B)}{\overline{P}_X(A)}, \tag{11}$$

and $Bel(B|A) = 1 - Pl(\bar{B}|A)$. Note that it comes down to the previous conditioning rule (7) with $\alpha_A(E) = 1$ if $E \cap A \neq \emptyset$, and 0 otherwise (an optimistic assignment, justified by the claim that A contains the actual value of x). Interestingly the conditioning rule for conjunctive random sets comes down to the previous conditioning rule (7) with $\alpha_A(E) = 1$ if $E \subseteq A$, and 0 otherwise, that could, in the belief function terminology, be written as $Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A)}$. It is known as the geometric rule of conditioning. Such a pessimistic weight reassignment can hardly be justified for disjunctive random sets.

4.2 Empirical Variance for Random Interval Data

Interval data sets provide a more concrete view of a random set. Again the distinction between the case where such intervals represent precise actual objects and when they express incomplete knowledge of precise ill-observed point values is crucial in computing a statistical parameter such as variance [7]. Consider a data set consisting of a bunch of intervals $\mathbb{D} = \{I_i = [\underline{a}_i, \overline{a}_i], i = 1, ...n\}$. The main question is: are we interested by the joint variation of the size and location of the intervals ? or are we interested in the variation of the underlying precise quantity as imperfectly accounted for by the variation of the interval data?

1. **Ontic interval data**: In this case we consider intervals are precise lumped entities. For instance, one may imagine the interval data set to contain sections of a piece of land according to coordinate x in the plane: $I_i = Y(x_i)$ for a multimapping Y, where $Y(x_i)$ is the extent of the piece of land at abscissa x_i, along coordinate y. The ontic view suggests the use of a scalar variance:
$$ScalVar(\mathbb{D}) = \frac{\sum_{i=1,\dots,n} d(M, I_i)^2}{n}, \tag{12}$$
where $M = [\sum_{i=1}^{n} \underline{a}_i/n, \sum_{i=1}^{n} \overline{a}_i/n]$ is the interval mean value, and d is a scalar distance between intervals (e.g. Euclidean distance between pairs of values representing the endpoints of the intervals, but more refined distances have been proposed [2]). $ScalVar(\mathbb{D})$ measures the variability of the intervals in $\mathbb{D}$, both in terms of location and width and evaluates the spatial regularity of the piece of land, varying coordinate x. This variance is 0 for a rectangular piece of land parallel to the coordinate axes.
2. **Epistemic interval data**: Under the epistemic view, each interval I_i stands for an ill-known precise value x_i that is the result of measuring a deterministic value x several times. Here, the measurement process is subject to randomness and is imprecise. Then we are more interested by sensitivity analysis describing what we know about the variance we would

have computed, had the data been precise. Then, we should compute the interval

$$EVar_1(\mathbb{D}) = \{var(\{x_i, i = 1, ...n\}) : x_i \in I_i, \forall i\}. \tag{13}$$

Computing this interval is a non-trivial task [25, 48]

3. **Epistemic interval random data** Alternatively one may consider that the quantity x that we wish to describe is intrinsically random. Each measurement process is an information source providing incomplete information on the variability of x. Then each interval I_i can be viewed as containing the support $SUPP(P_i)$ of an ill-known probability distribution P_i: then we get a wider variance interval than previously. It is defined by

$$EVar_2(\mathbb{D}) = \{var(\sum_{i=1}^{n} P_i/n) : SUPP(P_i) \subseteq I_i, \forall i = 1, ...n\} \tag{14}$$

and it is easy to see that $EVar_1(\mathbb{D}) \subset EVar_2(\mathbb{D})$.

In the extreme case of a single epistemic interval $(x, [a, b])$, if x is a deterministic ill-known quantity, it has a unique true value. Then $EVar_1([a, b]) = var(x) = 0$ (since even if ill-known, x is not supposed to vary: the set of variances of a bunch of Dirac functions is $\{0\}$). In the second case, x is tainted with variability, $var(x)$ is ill-known and lies in the interval $EVar_2([a, b]) = [0, v^*]$ where $v^* = \sup\{var(x), SUPP(P_x) \subseteq [a, b]\} = (b - a)^2$. The distinction between deterministic and stochastic variables known via intervals thus has important impact on the computation of dispersion indices, like variance.

Note that in the epistemic view, the scalar distance between intervals can be useful. It is then a kind of informational distance between pieces of knowledge, whose role can be similar to relative entropy for probability distributions. Namely one may use it in revision processes, for instance. Moreover one may be interested by the scalar variance of the imprecision of the intervals, or by an estimate of the actual variance of the underlying quantity, by computing the variance of say the mid-points of the intervals. Recently suggested scalar variances [44] between intervals come down a mixture of such a scalar variability estimation and the variance of imprecision.

5 Different Interpretations of a Fuzzy Set

A fuzzy set on a universe S is mathematically modelled by a mapping from S to a totally ordered set L that is usually the unit interval. As highlighted by Dubois and Prade [19], a membership function is an abstract object that needs to be interpreted in practical settings in order to be used meaningfully. They proposed three interpretations of membership grades in terms of degrees of similarity, of plausibility and preference. An early and important use of fuzzy sets, proposed by Zadeh [55] is the representation of symbolic categories on numerical universes. A linguistic variable is a variable that takes values on a set of linguistic terms modelled by fuzzy sets of the real line. In this case, degrees of membership express similarity or distance to prototypical values covered by a term.

As already acknowledged a long time ago, fuzzy sets, like sets, may have a conjunctive or a disjunctive reading [57, 53, 17]. In the conjunctive reading, ontic fuzzy sets represent objects originally construed as sets but for which a fuzzy representation is more expressive due to gradual boundaries. Degrees of membership evaluate to what extent components participate to the global entity. For instance, this is the case when modeling linguistic labels by convex fuzzy sets on a measurable scale, like *tall, medium-sized, short* achieving a fuzzy partition of the human height scale. In this case, the fuzzy sets have a conjunctive reading because they are understood as the set of heights compatible with a given label. Other examples of ontic fuzzy sets are non-Boolean classes stemming from a clustering process, fuzzy constraints representing preference, a fuzzy region in an image, a fuzzy rating profile according to various attributes. As a concrete example, consider the fuzzy set of languages more or less well spoken by a person.

In contrast, Zadeh [56] also proposed to interpret membership functions as possibility distributions, paving the way to a representation of incomplete information along a line followed thirty years earlier by Shackle [45]. In that case, a degree of membership refers to the idea of plausibility. A possibility distribution, denoted by π is the membership function of a fuzzy set of mutually exclusive values in S. A possibility distribution is supposedly attached to an ill-known quantity x. Namely $\pi(s) > 0$ expresses that s is a possible value of x, all the more plausible as $\pi(s)$ is greater. In particular it is assumed that $\pi(s) = 1$ for some value s, which is then considered as normal, totally unsurprizing. A possibility distribution thus extends the set-valued representation of incomplete information to account for degrees of plausibility. It is well-known that a possibility distribution π induces a possibility measure Π on 2^S such that $\Pi(A) = \sup_{s \in A} \pi(s)$ for all events A and a necessity measure $N(A) = 1 - \Pi(\bar{A})$ [16].

Now, if the information about a quantity x is expressed by means of a fuzzy set, the above distinction between the deterministic and the stochastic case is again at work. If x is deterministic, then this information must be interpreted in terms of "confidence sets" as follows. Let $E_\alpha = \{s, \pi(x) \geq \alpha\}$ be the α-cut of π:

For each $\alpha \in [0, 1]$, $x \in E_\alpha$ with probability greater than or equal to $1 - \alpha$.

If an expert provides this kind information, the word "probability" refers to subjective probability. Following Walley [52], $1 - \alpha$ is the maximal price at which this expert would buy the gamble that wins \$1 if the real value of x actually lies in E_α (the minimal selling price for this gamble is \$1). Note that there is no "real probability distribution" underlying π, but Dirac functions as x is deterministic. The consonance of the family of sets E_α makes sense if this is the opinion of a single expert who tends to be imprecise but self-consistent.

If x is stochastic then there are two possible ways of interpreting the possibility distribution π.

- Mathematically speaking, a possibility measure is a coherent upper probability [52], namely $\Pi(A) = \sup_{P \in \mathcal{P}_\pi} P(A)$ where $\mathcal{P}_\pi = \{P, \forall A, P(A) \leq \Pi(A)\}$. So, π encodes the set of probabilities $\mathcal{P}_\pi$ [18, 14]. This set is supposed to contain the real probability measure P_x that governs the variability of x. It is a set-based representation of a stochastic variable representing incomplete information about a frequentist probability. An expert providing distribution π claims that

  For each $\alpha \in [0, 1]$, the event $x \in E_\alpha$ has *objective* probability greater than or equal to $1 - \alpha$.

- Another option is to consider π as encoding a higher-order (subjective) possibility distribution on a set of objective probabilities. Namely, it can be understood as follows:

  For each $\alpha \in [0, 1]$, P_x has support in E_α with subjective probability greater than or equal to $1 - \alpha$.

  So the domain of π can be canonically extended to the set of probability measures on S as follows: $\pi(P) = \sup\{\alpha, P \text{ has support in } E_\alpha\}$. The possibility measure Π is a "second-order possibility" formally equivalent to those considered in [10]. It is so called, because it is a possibility distribution defined over a set of probability measures. The deterministic case is a special case of this framework, restricting probability measures to Dirac measures. It would be interesting to investigate the relationship between the set of probabilities $\mathcal{P}_\pi$ and the higher order possibility model.

The above setting does not make it clear where the objective probability distribution comes from, i.e. the underlying sample space. Moreover, it does not account for the measurement process of x. Namely, regardless of whether x is deterministic or stochastic, there may be a stochastic measurement process yielding with more or less accuracy information on the possible values of x. The setting of fuzzy random variables extends the above distinctions by taking the measurement process into account explicitly.

6 Various Notions of Random Fuzzy Sets

The history of fuzzy random variables is not simple as it was started by two separate groups with respectively epistemic and ontic views in mind. The first papers are those of Kwakernaak [33, 34] in the late seventies, clearly underlying an epistemic view of fuzzy sets, a line followed up by Kruse and Meyer

[32]. They view a fuzzy random variable as a (disjunctive) fuzzy set of classical random variables (those induced by selection functions compatible with the random fuzzy set). It represents what is known about the variability of the underlying ill-known random variable. These works can thus be viewed as extending the framework of Dempster's upper and lower probabilities based on the triple (Ω, P, X) to fuzzy set-valued mappings $\tilde{X}$, where $\tilde{X}(\omega)$ defines a possibility distribution restricting the possible values of $x(\omega)$. The degree of possibility that x is the random variable underlain by $(\Omega, P, \tilde{X})$ is

$$\pi(x) = \inf_{\omega \in \Omega} \mu_{\tilde{X}(\omega)}(x(\omega)) \tag{15}$$

For each level $\alpha \in (0,1], \tilde{X}_\alpha(\omega) = \{s \in S : \mu_{\tilde{X}(\omega)}(s) \geq \alpha\}$ is a multiple valued mapping such that $(\Omega, P, \tilde{X}_\alpha)$ is an epistemic random set according to Dempster framework. Kruse and Meyer [32] clearly define the variance of a fuzzy random variable as a fuzzy set of positive reals induced by applying the extension principle to the variance formula. Likewise, the probability of an event becomes restricted by a fuzzy interval in the real line [1]. The evidence theory counterpart of this view deals with belief functions having fuzzy focal elements [54]. An alternative epistemic view of fuzzy random variables was more recently proposed in the spirit of Walley [52], in terms of a convex set of probabilites induced on S[8].

In contrast, the line initiated in the mid-1980's by Puri and Ralescu [43] is in agreement with conjunctive random set theory. A fuzzy random variable is then viewed as a random conjunctive fuzzy set, i.e. a classical random variable ranging in a set of (membership) functions. This line of research has been considerably extended so as to adapt classical statistical methods to functional data [5, 27]. The main issue is to define a space of functions equipped with a suitable metric structure [13, 51]. In this theory of random fuzzy sets, a scalar distance between fuzzy sets is instrumental when defining variance viewed as a mean squared deviation from the fuzzy mean value [28], in the spirit of Fréchet. A scalar variance can be established on this basis and it reflects the variability *of membership functions*. It makes sense if for instance, membership functions are models of linguistic terms and some "term variability" must be evaluated given a set of responses provided by a set of people in natural language. See [7] for an extensive comparison of the three views of fuzzy random variables.

The ontic view is advocated by Colubi *et al.* [6] in the statistical analysis of linguistic data. The authors argue that they are interested in the statistics of perceptions. One of their experiments deals with the visual perception of the length of a line segment expressed on fuzzy scale using a linguistic label among *very small, small, medium, large, very large*. The alleged goal is to predict the category that a person considers correct for the segment. The precise length of the segment exists but it is irrelevant for the classification goal. They agree that to predict the real length from the fuzzy perceptions requires a different approach.

The case of Likert scaling is more problematic. This is a method of ascribing quantitative values to qualitative data, to make them amenable to statistical analysis. For instance, an ordered set of linguistic labels referring to some abstract concept (like beauty) is encoded by successive integers. A typical scale might be *strongly agree, agree, not sure/undecided, disagree, strongly disagree*. Opinions are collected on such a scale and a mean figure for all the responses is computed at the end of the evaluation or survey. A number of authors have proposed to model such linguistic terms by means of a predefined fuzzy partition made of fuzzy intervals (trapezoids) on a real interval. In some other approaches the format of the fuzzy response can be any fuzzy interval. The idea is to cope with the arbitrariness of encoding qualitative value by precise numbers. In that case the result of an opinion poll is clearly a random fuzzy set.

However this kind of approach is not convincing from a measurement point of view[15]. First, it is not clear why the underlying real interval can be equipped with addition at all. It is rather an ordinal scale, and trapezoidal fuzzy sets then make no sense. Next, this continuous scale is totally fictitious and it is patent that the real data are the linguistic terms provided by people: there is no underlying real value behind such linguistic terms. If the response has a free format (whereby any fuzzy interval can do), one may again see this fuzzy response as being the evaluation in itself. The latter point would plea for an ontic view of the random fuzzy sets. However the arbitrariness of the numerical encoding casts doubts on the cogency of the sophisticated functional analysis framework needed to apply fuzzy random set methods. It may be that ordinal statistical methods devoted to finite qualitative scales would be more appropriate in this case.

7 Epistemic vs. Ontic Interval Data Processing

Consider a set of bidimensional interval data $\mathbb{D} = \{(x_i, Y_i = [\underline{y}_i, \overline{y}_i]), i = 1, ...n\}$ or its fuzzy counterpart (if the Y_i's become fuzzy sets). The issue of devising an extension of data processing methods to such a situation has been studied in many papers in the last 20 years or so. But it seems that the question how the reading of the set-valued data has impact on the chosen method is seldom discussed. Here we provide some hints on this issue, restricting ourselves to linear regression and some of its fuzzy extensions.

A first approach that is widely known is Diamond's fuzzy least squares method [12]. It is based on a scalar distance between set-valued data. The problem is to find a best fit interval model of the form $y = A^*x + B^*$, where intervals A^*, B^* minimize $\sum_{i=1}^{n} d(Ax_i + B, Y_i)^2$, typically a sum of squares of differences between upper and lower bounds of intervals. The fuzzy least squares regression is similar but it presupposes the $\tilde{Y}_i'$s are triangular fuzzy intervals $(y_i^m; y_i^-, y_i^+)$, with modal value y_i^m and support $[y_i^-, y_i^+]$.

Diamond proposes to work with a scalar distance of the form $d^2(\tilde{A}, \tilde{B}) = (a^m - b^m)^2 + (a^- - b^-)^2 + (a^+ - b^+)^2$ making the space of triangular fuzzy intervals complete. The problem is then to find a best fit fuzzy interval model $\tilde{Y} = \tilde{A}^* x + \tilde{B}^*$, where fuzzy triangular intervals $\tilde{A}^*, \tilde{B}^*$ minimize a squared error $\sum_{i=1}^{n} d^2(\tilde{A}x_i + \tilde{B}, \tilde{Y}_i)$. Some comments are in order:

- This approach treats fuzzy data as ontic entities.
- If the (fuzzy) interval data-set is epistemic, we get a linear description of the trend of the knowledge as x increases.
- This approach does not correspond to studying the impact of data uncertainty on the result of regression.

Many variants of this method, based on conjunctive fuzzy random sets and scalar distances exist (see [24] for a recent one) including extensions to fuzzy-valued inputs [26]. These approaches all adopt the ontic view.

Another classical approach was proposed by Tanaka *et al.* in the early 1980's (see [50] for an overview). One way of posing the interval regression problem is to find a set-valued function $Y(x)$ (generally again of the form of an interval-valued linear function $Y(x) = Ax + B$) with maximal informative content such that $Y_i \subset Y(x_i), i = 1, ...n$. Some comments help situate this method:

- It does not presuppose an ontic or epistemic reading of the data. If data are ontic, the result models an interval-valued phenomenon. If epistemic, it tries to cover both the evolution of the variable y and the evolution of knowledge of this phenomenon.
- It does not clearly extend the basic concepts of linear regression.

Both approaches rely on the interval extension of a linear model $y(x) = ax+b$. But, in the epistemic reading, this choice imposes unnatural constraints on the relation between the epistemic output $Y(x)$ and the objective input x (e.g., $Y(x)$ becomes wider as x increases). The fact that the real phenomenon is affine does not imply that the knowledge about it in each point should be also of the form $Y(x) = Ax + B$. In an ontic reading, one may wish to interpolate the interval data more closely (see Boukezzoula *et al.* [3] for improvements of Tanaka's methods that cope with such defects).

Another view of interval regression, that has a clear epistemic flavor uses possibility theory to define a kind of quantile regression. Even when applied to precise data sets it gives an epistemic interval-valued representation of objective data, likely to contain the actual model [42]. The idea is to find, for each input value x, a confidence interval containing $y(x)$ with confidence level $1-\alpha$. This is done via probability possibility transformations [22]. Varying α leads to a bunch of nested intervals that can be modelled by fuzzy intervals faithful to the dispersion of the y_i's in the vicinity of each input data x_i.

The last approach we can think of is sensitivity analysis yielding all regression results one would obtain from all precise datasets $\mathbf{d}$ consistent with $\mathbb{D}$. Strangely enough this technique is seldom proposed. The aim is to find the range of results one would have obtained with linear regression, had the data been precise. Formally it can be posed as follows: Find

$$Y(x) = \{\hat{a}(\mathbf{d})x + \hat{b}(\mathbf{d}) : \mathbf{d} \in \mathbb{D}\}$$
$$=\{\hat{a}x + \hat{b}, \forall \hat{a}, \hat{b} \text{ that minimize } \textstyle\sum_{i=1}^{n}(ax_i + b - y_i)^2, \forall y_i \in [\underline{y}_i, \overline{y}_i], i = 1, ...n\}$$

It is clear that the envelope of the results is a set-valued function $Y(x)$ that has little chance of being defined by affine upper and lower bounds. This approach, which can genuinely be called epistemic regression has been recently applied to kriging in geostatistics [36, 37].

8 Conclusion

This position paper has argued that the use of set-valued and fuzzy mathematics in information processing tasks gives the opportunity to reason about knowledge, an issue not so popular in data-driven studies. However, one should distinguish between genuine set-valued problems where sets stand for existing entities and epistemic data analysis problems where sets represent incomplete information. This distinction impacts the very way new problems can be posed so as to be meaningful in practice. Adding knowledge representation and reasoning to the modeling paradigm seems to be a good way to reconcile Artificial Intelligence and numerical engineering methods.

Strangely enough fuzzy set-based information processing techniques gathered under the Soft Computing flag are not set-valued methods, as they aim most of the time at computing standard numerical functions using fuzzy rules and neural networks, exploiting stochastic metaheuristics to optimise the fit. A fuzzy system is then seldom viewed as an epistemic fuzzy set of systems. Adopting the latter view could lead to fruitful developments of fuzzy sets methods in a direction not yet much considered in the engineering sciences, beyond rehashing good old fuzzy rule-based systems further.

References

1. Baudrit C, Couso I, Dubois D (2007) Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *Int J Approx Reas* 45:82–105
2. Bertoluzza ASC, Salas A, Corral N (1995) On a new class of distances between fuzzy numbers. *Mathware and Soft Comp* 2:71–84
3. Boukezzoula R, Galichet S, Bisserier A (2011) A Midpoint-Radius approach to regression with interval data. *Int. J. Approx. Reasoning* 52:1257–1271
4. Bouyssou D, Dubois D, Pirlot M, Prade H, eds. (2009) *Decision-making Process — Concepts and Methods.* ISTE London & Wiley, New-York
5. Colubi A (2009) Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. *Fuzzy Sets Syst* 160(3):344–356

6. Colubi A, González-Rodríguez G, Gil MA, Trutschnig W (2011) Nonparametric criteria for supervised classification of fuzzy data. *Int. J. Approx Reas* 52:1272–1282
7. Couso I, Dubois D (2009) On the Variability of the Concept of Variance for Fuzzy Random Variables. *IEEE Trans Fuzzy Syst* 17:1070–1080
8. Couso I, Sánchez L (2011) Upper and lower probabilities induced by a fuzzy random variable. *Fuzzy Sets Syst* 165:1–23
9. De Campos LM, Lamata MT, Moral S (1990) The concept of conditional fuzzy measure. *Int. J. of Intell Syst* 5:237–246
10. De Cooman G, Walley P (2002) An imprecise hierarchical model for behaviour under uncertainty. *Theory and Decision* 52:327–374
11. Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat* 38:325–339
12. Diamond P (1988) Fuzzy least squares. *Inform Sci* 46:141–157
13. Diamond P, Kloeden P (1994) Metric spaces of fuzzy sets. World Scientific, Singapore
14. Dubois D (2006) Possibility theory and statistical reasoning. *Comp Stat & Data Anal* 51:47–69
15. Dubois D (2011) The role of fuzzy sets in decision sciences: Old techniques and new directions. *Fuzzy Sets Syst* 184:3–28
16. Dubois D, Prade H (1988) *Possibility Theory.* Plenum Press, New York
17. Dubois D, Prade H (1988) Incomplete conjunctive information. *Comp & Math Appl* 15:797–810
18. Dubois D, Prade H (1992) When upper probabilities are possibility measures. *Fuzzy Sets Syst* 49:65–74
19. Dubois D, Prade H (1997) The three semantics of fuzzy sets. *Fuzzy Sets Syst* 90:141–150
20. Dubois D, Prade H (2009) Formal representations of uncertainty. Chap. 3 in [4], 85–156
21. Dubois D, Prade H (2012) Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets Syst* 192:3–24
22. Dubois D, Foulloy L, Mauris G, Prade H (2004) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing* 10:273–297
23. Fagin R, Halpern JY (1991) A new approach to updating beliefs. In: Bonissone PP, Henrion M, Kanal LN, Lemmer JF (eds.) *Uncertainty in Articial Intelligence (UAI'91)*, 347–374. Elsevier, New York
24. Ferraro MB, Coppi R, González-Rodríguez G, Colubi A (2010) A linear regression model for imprecise response. *Int J Approx Reas* 51:759–770
25. Ferson S, Ginzburg L, Kreinovich V, Longpre L, Aviles M (2002) Computing variance for interval data is NP-hard. *ACM SIGACT News* 33:108–118
26. González-Rodríguez G, Blanco A, Colubi A, Lubiano MA (2009) Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets and Systems* 160(3):357–370
27. González-Rodríguez G, Colubi A, Gil MA (2012) Fuzzy data treated as functional data. A one-way ANOVA test approach. *Comp Stat and Data Anal* 56(4):943–955
28. Körner R (1997) On the variance of fuzzy random variables. *Fuzzy Sets Syst* 92:83–93
29. Halpern JY, Fagin R, Moses Y, Vardi MY (2003) *Reasoning About Knowledge.* MIT Press, Cambridge
30. Herzig A, Lang J, Marquis P (2003) Action representation and partially observable planning using epistemic logic. *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI-03)*, 1067–1072. Morgan Kaufmann, San Francisco
31. Kendall DG (1974) Foundations of a theory of random sets. In: Harding EF, Kendall DG, eds. *Stochastic geometry*, 322–376 J. Wiley & Sons, New York
32. Kruse R, Meyer K (1987) *Statistics with Vague Data.* D. Reidel, Dordrecht
33. Kwakernaak H (1978) Fuzzy random variables — I. definitions and theorems. *Inform Sci* 15:1–29

34. Kwakernaak H (1979) Fuzzy random variables — II. Algorithms and examples for the discrete case. *Inform Sci* 17:253–278
35. Lindley DV (1982) Scoring rules and the inevitability of probability. *Int Statist Rev* 50:1–26
36. Loquin K, Dubois D (2010) Kriging and epistemic uncertainty: a critical discussion. In: Jeansoulin R *et al.* (eds). *Methods for Handling Imperfect Spatial Information*, 269–305. Springer, Berlin Heidelberg New York
37. Loquin K, Dubois D (2010) Kriging with ill-known variogram and data. *Scalable Uncertainty Management (SUM 2010)*, LNAI 6379:219–235. Springer, Berlin Heidelberg New York
38. Matheron G (1975) *Random Sets and Integral Geometry.* J. Wiley & Sons, New York, NY, USA
39. Moore R (1979) Methods and Applications of Interval Analysis. *SIAM Studies in Applied Mathematics.* SIAM, Philadelphia
40. Nguyen HT (1978) On random sets and belief functions. *J Math Anal Appl* 65:531–542
41. Pichon F, Dubois D, Denoeux T (2012) Relevance and truthfulness in information correction and fusion. *Int J Approx Reas* 53:159–175
42. Prade H, Serrurier M (2011) Maximum-likelihood principle for possibility distributions viewed as families of probabilities. *Proc. IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE 2011)*, 2987–2993. IEEE Press, Piscataway
43. Puri M, Ralescu D (1986) Fuzzy random variables. *J Math Anal Appl* 114:409–422
44. Ramos-Guajardo AB, Lubiano MA (2012) K-sample tests for equality of variances of random fuzzy sets. *Comp Stat & Data Anal* 56:956–966
45. Shackle GLS (1961) *Decision, Order and Time in Human Affairs (2nd edition).* Cambridge University Press
46. Shafer G (1976) *A Mathematical Theory of Evidence.* Princeton University Press
47. Shafer G, Tversky A (1985) Languages and designs for probability. *Cogn Sci* 9:309–339
48. Spadoni M, Stefanini L (2011) Computing the variance of interval and fuzzy data. *Fuzzy Sets Syst* 165:24–36
49. Smets P (1997) The normative representation of quantified beliefs by belief functions. *Artif Intell* 92:229–242
50. Tanaka H, Guo P (1999) *Possibilistic Data Analysis for Operations Research.* Physica-Verlag, Heidelberg
51. Trutschnig W, González-Rodríguez G, Colubi A, Gil MA (2009) A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. *Inform Sci* 179:3964–3972
52. Walley P (1991) *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall
53. Yager RR (1987) Set-based representations of conjunctive and disjunctive knowledge. *Inform Sci* 41:1–22
54. Yen J (1990) Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Trans Syst Man and Cybern* 20:559–569
55. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning, part I. *Inform Sci* 8:199–249
56. Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:1–28
57. Zadeh LA (1978) PRUF — a meaning representation language for natural languages. *Int J Man-Mach Stud* 10:395–460