View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Scientific Publications of the University of Toulouse II Le Mirail

# Statistical reasoning with set-valued information : Ontic vs. epistemic views

Inès Couso, Didier Dubois

**To cite this version:**

Inès Couso, Didier Dubois. Statistical reasoning with set-valued information : Ontic vs. epistemic views. International Journal of Approximate Reasoning, Elsevier, 2014, vol. 55 (n 7), pp. 1502-1518. <10.1016/j.ijar.2013.07.002>. <hal-01153809>

**HAL Id: hal-01153809**
**https://hal.archives-ouvertes.fr/hal-01153809**

Submitted on 20 May 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 13206

**To link to this article** : DOI:10.1016/j.ijar.2013.07.002
URL : http://dx.doi.org/10.1016/j.ijar.2013.07.002

**To cite this version** : Couso, Inès and Dubois, Didier *Statistical reasoning with set-valued information : Ontic vs. epistemic views.* (2014) International Journal of Approximate Reasonning, vol. 55 (n° 7). pp. 1502-1518. ISSN 0888-613X

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

# Statistical reasoning with set-valued information: Ontic vs. epistemic views

Inés Couso [a], Didier Dubois [b]

[a] *Department of Statistics, Universidad de Oviedo, Spain*
[b] *IRIT, CNRS Université de Toulouse, France*

A B S T R A C T

In information processing tasks, sets may have a conjunctive or a disjunctive reading. In the conjunctive reading, a set represents an object of interest and its elements are subparts of the object, forming a composite description. In the disjunctive reading, a set contains mutually exclusive elements and refers to the representation of incomplete knowledge. It does not model an actual object or quantity, but partial information about an underlying object or a precise quantity. This distinction between what we call ontic vs. epistemic sets remains valid for fuzzy sets, whose membership functions, in the disjunctive reading are possibility distributions, over deterministic or random values. This paper examines the impact of this distinction in statistics. We show its importance because there is a risk of misusing basic notions and tools, such as conditioning, distance between sets, variance, regression, etc. when data are set-valued. We discuss several examples where the ontic and epistemic points of view yield different approaches to these concepts.

*Keywords:*
Incomplete information
Random sets
Fuzzy sets
Evidence theory
Imprecise probability
Possibility theory

## 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 a simplified abstraction 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 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 [9]. 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 [42]. Representing knowledge requires a logical language and this approach has been mainly developed in the framework of classical or modal logic, due to the long philosophical tradition in this area. Contrary to the numerical modeling tradition, such knowledge-based models are most of the time tainted with incompleteness: a set of logical formulae, representing an agent's beliefs is seldom complete, that is, cannot establish the truth or falsity of any proposition. This concern for incomplete information in Artificial Intelligence has strongly affected the development of new uncertainty theories [32], and has led to a critique of the Bayesian stance viewing probability theory as a unique framework for the representation of belief that mimics the probabilistic account of variability.

These developments question traditional views of modeling as representing reality independently of human perception and reasoning. They suggest a different approach where mathematical models 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 position paper[1] 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. The outline of the paper is as follows. In Section 2, we discuss the use of sets for the representation of epistemic states as opposed to the representation of objective entities. Then in Section 3 we draw the consequences of this discussion in the theory of random sets, laying bare three approaches relying on the same mathematical tool. In Section 4, we show that the distinction drawn between epistemic and ontic random sets affects the practical relevance of formal definitions one can pose in the random set setting. It is shown that notions of conditioning, independence and variance differ according to the adopted point of view. The consequences of this distinction in the way interval regression problems can be posed are briefly discussed in Section 5.2. Finally, Section 6 carries the distinction between ontic and epistemic sets over to fuzzy sets, and, more briefly, to random fuzzy sets.

## 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 [33] 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 time points 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. Moreover this set is likely to change by acquiring more information. The use of sets representing imprecise values can be found for instance in interval analysis [54]. Another example is the set of models of a proposition in logic, or a propositional knowledge base: only one of them reflects the real situation; this is reflected by the DNF form of a proposition, i.e., a disjunction of its models, each of which is a maximal conjunction of literals.

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 [43]. 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 [77]. This distinction between conjunctive and disjunctive sets was already made by Zadeh [78] 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). The study of incomplete

[1] An expanded version of a first draft (by the second author) that is part of the COST702 final report published by Springer as an edited volume "Towards Advanced Data Analysis by Combining Soft Computing and Statistics", in: C. Borgelt, et al. (Eds.), Studies in Fuzziness and Soft Computing, vol. 285, 2013.

conjunctive information (whose representation requires a disjunctive set of conjunctive sets) was carried out quite early and can be found in papers by Yager [74] and Dubois and Prade [29]. More recently Denoeux et al. [21] have proposed an approach to formalize uncertain conjunctive information using belief functions.

An epistemic set $(x, E)$ does not necessarily account 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 at a specific location is a stochastic variable that 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 $\Omega$, and $x$ can be a random variable taking values on $S$, but the probability distribution induced by $x$ is unknown.[2] All that is known is that $x(\omega) \in E$, for all $\omega \in \Omega$. It implies $P_x(E) = 1$ where $P_x$ is the probability measure of $x$. Then, $E$ represents the family $\mathcal{P}_E$ of objective probability measures on $\Omega$ such that $P(\{\omega\colon x(\omega) \in E\}) = 1$, 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. This approach was until recently often adopted as natural if $x$ is a random variable. In case $x$ is an ill-known deterministic value, Bayesians [50] propose to use a subjective probability $P_x^b$ in place of set $E$. Then, the occurrence of $x$ is not a matter of repetitions, and the degree of probability is made sense of via a betting interpretation on a single occurrence of event $A$: $P_x^b(A)$ 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 [32] 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 [5]), 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 $(\Omega, P)$ is available,[3] but each realization of the random variable is represented as a set. This case covers two situations:

1. **Random conjunctive sets**: The random variable $X(\omega)$ is multivalued 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 $\omega$. Or $X(\omega)$ is an ill-known area of interest in some spatial domain, and $\omega$ 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 [45], Matheron [53]). 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 realizations 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 (often called selections) $\{x\colon\ \Omega \to S, \forall \omega, x(\omega) \in X(\omega)\} = \{x \in X\}$ for short. In other words, the triple $(\Omega, P, X)$ is an epistemic model of the random variable $x$ and the probability $p_X(E) = \sum_{\omega\colon\ X(\omega)=E} p(\omega)$ represents the proportion of the population in $\Omega$ for which all that is known is that $x(\omega) \in E$.
   This is the approach of Dempster [19] to imprecise probabilities. He uses this setting to account for a parametric probabilistic model $P_\theta$ 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 $X(\omega) = \{(\theta, u)\colon\ \phi(u, \theta) = \omega\}$. It is clear that, for each observation $u$, $X_u(\omega) = \{\theta\colon\ \phi(u, \theta) = \omega\}$ is an epistemic set restricting the actual (deterministic) value $\theta$, and the probability distribution of $\phi$ generates an epistemic random set on the parameter space $\Theta$.

Shafer [62] 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

[2] We will use lower case for both deterministic and random point-valued quantities, in contrast with set-valued deterministic/random objects, that will be denoted by capitals.

[3] In this paper, we assume $\Omega$ is finite to avoid mathematical difficulties. The probability measure $P$ will be assumed to be defined over the discrete $\sigma$-field, $\wp(\Omega)$.

to the empty information $S$. This view of probability was popular until the end of the 18th century (see [57] 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 $u$ [63].

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 $S = \{\text{English}, \text{French}, \text{Spanish}\}$, one may compute the probability that someone speaks English by summing the proportions of people in $\Omega$ 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 $\omega$ 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 $\Omega$ lies in $A$, namely lower and upper probabilities proposed by Dempster [19]:

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

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

such that $\underline{P}_X(A) = 1 - \overline{P}_X(A^c)$, where $A^c$ is the complement of $A$. Note that the set of probabilities $\mathcal{P}_X$ on $S$ induced by this process is finite: since $\Omega$ 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 $\hat{\mathcal{P}}_X = \{P; \forall A \in S, P(A) \geqslant \underline{P}_X(A)\}$. It is well-known that probability measures in this convex set are of the form

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

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 [10], 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:

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

which lies in $\hat{\mathcal{P}}_X$. However, all we know about $P(A|\omega)$ 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

$$\underline{P}_X(A) = \inf\left\{\sum_{\omega\in\Omega} P(A|\omega)p(\omega)\colon\ P\big(X(\omega)|\omega\big) = 1, \forall\omega \in \Omega\right\} \tag{4}$$

$$= \inf_{\{P_E\colon E\subseteq S\}} \sum_{E\subseteq S} p_X(E)P_E(A). \tag{5}$$

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 [69]. The mathematical similarity between belief functions and random sets was quite early pointed out by Nguyen [55]. 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: conditioning and independence

The reader may consider that the difference between the three above interpretations of random sets is just a philosophical issue, that does not impact on the definition of basic notions that can be studied in each case. 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 here: conditioning and independence.

### 4.1. *Conditioning random sets*

Given a random set in the form of a probability distribution on the power set of $S$, and a non-impossible 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 realizations $X(\omega)$ 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:

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

**Example 1.** Let $S$ be a set of spoken languages including $c =$ Chinese, $e =$ English, $f =$ French and $s =$ Spanish. Let us consider the sets $C = \{e, f, s\}$ and $A = S \setminus \{c\}$. Then $p_X(C|A)$ denotes the proportion of people that speak English, French and Spanish (and nothing else), among those who cannot speak Chinese.

Eq. (6) can be generalized by considering a pair of arbitrary families of subsets of $S$, $\mathcal{C}$ and $\mathcal{A}$, and writing

$$P_X(\mathcal{C}|\mathcal{A}) = \frac{P_X(\mathcal{C} \cap \mathcal{A})}{P_X(\mathcal{A})} = \frac{\sum_{B\in\mathcal{C}\cap\mathcal{A}} p_X(B)}{\sum_{A\in\mathcal{A}} p_X(A)}. \tag{7}$$

Eq. (7) is nothing else but the classical rule of conditioning, where the images of the random set $X$, which are subsets of $S$, are seen as "elements" of the actual "universe", and the families of subsets of $S$ are the actual "events". In fact, in Eq. (6), we should write $p_X(C|\mathcal{A})$ where $\mathcal{A} = \{B\colon\ B \subseteq A\}$.

**Example 2.** Let $S$ be again the set of spoken languages considered in Example 1 and let us consider the families of sets $\mathcal{C} = \{C\colon\ C \ni e\}$ and $\mathcal{A} = \{A\colon\ A \not\ni c\}$. Then, $p_X(\mathcal{C}|\mathcal{A}) = \frac{\sum_{C\in\mathcal{C}\cap\mathcal{A}} p_X(C)}{\sum_{A\in\mathcal{A}} p_X(A)}$ represents the proportion of people that can speak (at least) English, among those who do not speak Chinese.

*Conditioning an ill-known random variable* Suppose the epistemic random set $X(\omega)$ relies on a population $\Omega$, and is represented by a convex set $\hat{\mathcal{P}}_X$ of probabilities $P_x$ 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 recomputed by restricting $\Omega$ to the subset $\Omega_A = \{\omega\colon\ x(\omega) \in A\}$ of population $\Omega$. However, because $x(\omega)$ is only known to lie in $X(\omega)$, the set $\Omega_A$ is itself ill-known.

Suppose we know for each focal set $E$, the proportion $\alpha_A(E)$ of the population (for which all we know is $x(\omega) \in E$) that lies inside $\Omega_A$. Namely:

- If $A$ and $E$ are disjoint sets, $\alpha_A(E)$ will be equal to zero.
- Moreover, if $E \subseteq A$, then $\alpha_A(E) = 1$ necessarily.
- In the rest of the cases ($E \cap A \neq \emptyset$ and $E \cap A^c \neq \emptyset$), $\alpha_A(E)$ may take any value between 0 and 1.

We can obtain $\alpha_A$ if, for each non-empty subset $E$, we know the probability distribution $p_E$ such that $p_E(s) = P(x(\omega) = s|X(\omega) = E)$ sharing the mass $p_X(E)$ defined in the previous Section 3. It is clear that we can define $\alpha_{\{s\}}(E) = p_E(\{s\})$, so that $\sum_{s\in E} \alpha_{\{s\}}(E) = 1$, $\forall E \in \mathcal{F}$ and:

$$\alpha_A(E) = \sum_{i:\ s\in A} \alpha_{\{s\}}(E) = P_E(A), \quad \forall A \subseteq S,\ \forall E \in \mathcal{F}.$$

It defines a set $\mathbb{A} = \{\alpha_A(E)\colon\ A \subseteq S, E \in \mathcal{F}\}$ of coefficients taking values on the unit interval $[0, 1]$. Each set $\mathbb{A}$ determines a probability measure, $P_x^\alpha \in \hat{\mathcal{P}}_X$ such that:

$$p_x^\alpha(s) = \sum_{E\in\mathcal{F}} \alpha_{\{s\}}(E) \cdot p_X(E), \quad \forall s \in S. \tag{8}$$

If the set of coefficients $\{\alpha_A(E), E \in \mathcal{F}\}$ is fixed, it is possible to condition $p_X$ by restricting it to subsets of $A$ as follows [17]:

$$p_X^{\alpha_A}(F|A) = \begin{cases} \frac{\sum_{F=E\cap A} \alpha_A(E) p_X(E)}{\sum_{E\cap A\neq\emptyset} \alpha_A(E) p_X(E)} & \text{if } F \subseteq A; \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The idea is that the probability assigned to a subset $F$ of $A$, when restricting to the population $\Omega_A$ is obtained by assigning to it all fractions $\alpha_A(E)p_X(E)$ of the population in $\Omega$ for which all we know is that $x(\omega) \in E$ and that we

**Table 1**
Joint distribution of the pair of random sets $(X_1, X_2)$.

| $X_1 \backslash X_2$ | $\{r\}$ | $\{w\}$ | $\{r, w\}$ |
|---|---|---|---|
| $\{r\}$ | 1/6 | 1/6 | 0 |
| $\{w\}$ | 1/6 | 1/6 | 0 |
| $\{r, w\}$ | 0 | 0 | 1/3 |

know to actually lie in $F = A \cap E$. It gives birth to plausibility and belief functions $\overline{P}_X^{\alpha_A}(B|A) = \sum_{F \cap B \neq \emptyset} p_X^{\alpha_A}(F|A)$ and $\underline{P}_X^{\alpha_A}(B|A) = \sum_{F \subseteq B} p_X^{\alpha_A}(F|A)$ over $A$. Varying the vector $(\alpha_A(E)_{E \subseteq S})$ leads to conjugate upper and lower set-functions as follows [17]:

$$\overline{P}_X(B|A) = \sup_{\alpha_A} \overline{P}_X^{\alpha_A}(B|A); \qquad \underline{P}_X(B|A) = \inf_{\alpha_A} \underline{P}_X^{\alpha_A}(B|A). \tag{10}$$

These bounds are attained by the following choices of $\alpha_A$ vectors, where $B \subseteq A$:

- Upper bound $\overline{P}_X(B|A)$: $\alpha_A(E) = 1$ if $E \cap B \neq \emptyset$ or $E \subseteq A$, 0 otherwise.
- Lower bound $\underline{P}_X(B|A)$: $\alpha_A(E) = 1$ if $E \cap A \subseteq B$ or $E \subseteq A$, 0 otherwise.

In fact, it has been proved that the same bounds can be obtained by applying Bayesian conditioning to all probabilities in $\mathcal{P}_X$ with $P_x(A) > 0$. They are upper and lower conditional probabilities that take an attractive closed form [17,34]:

$$\overline{P}_X(B|A) = \sup\{P_x(B|A)\colon\ P_x \in \mathcal{P}_X\} = \frac{\overline{P}_X(B \cap A)}{\overline{P}_X(B \cap A) + \underline{P}_X(B^c \cap A)}, \tag{11}$$

$$\underline{P}_X(B|A) = \inf\{P_x(B|A)\colon\ P_x \in \mathcal{P}_X\} = \frac{\underline{P}_X(B \cap A)}{\underline{P}_X(B \cap A) + \overline{P}_X(B^c \cap A)}, \tag{12}$$

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

In fact this is not surprizing since each vector $\alpha_A$ corresponds to a subset of probability measures $P_x^{\alpha} \in \mathcal{P}_X$ obtained from all sets $\mathbb{A}$ of coefficients containing $\{\alpha_A(E)\colon\ E \in \mathcal{F}\}$ and each $P_x \in \mathcal{P}_X$ is generated by some set $\mathbb{A}$: more precisely $P_x^{\alpha}(B) = \sum_{E \subseteq S} p_X(E)\alpha_B(E)$ due to Eqs. (3) and (8). Noticeably, $\overline{P}_X(B|A)$ and $\underline{P}_X(B|A)$ are still plausibility and belief functions, as proved in [40,56], so that Eqs. (9) and (10) justify this form of conditioning (familiar in imprecise probability theory [73]) in the setting of belief functions.

The following example is a variant of Example 6 in [11].

**Example 3.** Suppose that we have three urns. The first one has 3 balls: one white, one red and one unpainted. The second urn has two balls: one red and one white. The third urn has two unpainted balls. We randomly select one ball from the first urn. If it is colored, then we randomly select a second ball from the second urn. Else if it is unpainted, we select the second ball from the third urn. Once the two balls have been selected, they will be painted red or white according to an unknown procedure. The information about the final color of the pair of randomly selected balls can be represented by means of a random set $X = X_1 \times X_2$ taking values on the product space $\Omega \times \Omega = \{r, w\} \times \{r, w\}$, where $X_1$ and $X_2$ respectively denote our incomplete information about the final color of the first and the second ball, respectively. $X_1$ and $X_2$ are set-valued functions, each one of them with the possible outcomes $\{r\}$ (red color) $\{w\}$ (white color) and $\{r, w\}$ (unknown color). According to above process of selection, the probability of appearance of each pair of outcomes is provided in Table 1.

The actual pair of colors of both (randomly selected) balls can be represented by means of a random vector $x = (x_1, x_2)$. All we know about it is that its values belong to the (set-valued) images of $X = X_1 \times X_2$, or, in other words, that the random vector $x$ belongs to the family of selections of $X$, $S(X)$. Let us now consider, for instance the events $A = \{w\} \times \Omega$ and $B = \Omega \times \{w\}$. The conditional probability $p_x(B|A)$ denotes the probability that the color of second ball $(x_2)$ is also white, if we know that the first one $(x_1)$ was white. Our knowledge about this value is determined by the pair of bounds

$$\overline{P}_X(B|A) = \sup_{y \in S(X)} p_y(B|A) = \frac{1/6 + 1/3}{1/6 + (1/6 + 1/3)} = 3/4$$

and

$$\underline{P}_X(B|A) = \inf_{y \in S(X)} p_y(B|A) = \frac{1/6}{(1/6 + 1/3) + 1/6} = 1/4.$$

*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 [19]). 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 normalizes the conditional mass function, dividing by $\sum_{E\cap A\neq\emptyset} m(E) = Pl(A)$. It leads to the conditioning rule

$$Pl(B|A) = \frac{Pl(A\cap B)}{Pl(A)} = \frac{\overline{P}_X(A\cap B)}{\overline{P}_X(A)}, \tag{13}$$

and $Bel(B|A) = 1 - Pl(B^c|A)$. Note that it comes down to the previous conditioning rule (9) 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 $x$).

**Example 4.** Let us suppose that we have the following incomplete information about an unknown value $x$: It belongs to the interval $E_1 = [25, 35]$ with probability greater than or equal to 0.5. Furthermore, we are completely sure that it belongs to the (less precise) interval $E_2 = [20, 40]$. This information can be represented by means of a consonant mass function $m$ assigning the following masses to the above intervals (the focal elements):

$$m\big([25,35]\big) = 0.5 \quad \text{and} \quad m\big([20,40]\big) = 0.5.$$

According to this information the degree of belief associated to the set $B = [30, 35]$, $Bel(B) = \sum_{C\subseteq B} m(C) = 0$ and $Pl(B) = 1$. Now, in addition, we are told that $x$ belongs, for sure, to the interval of values $A = [30, 40]$. In order to update our information, we transfer the masses $m(E_1)$ and $m(E_2)$ respectively to $E_1 \cap A$ and $E_2 \cap A$. The updated information is therefore represented by means of a new consonant mass assignment $m'$ defined as $m'([30,35]\cap[30,40]) = m'([30,35]) = 0.5$, and $m'([30,40]) = m'([20,40]\cap[30,40]) = 0.5$. Let us calculate the conditional belief $Bel(B|A)$. According to Dempster's rule of combination we get:

$$\begin{aligned} Bel(B|A) &= 1 - Pl\big(B^c|A\big) = 1 - \frac{Pl(B^c\cap A)}{Pl(A)} \\ &= 1 - \frac{Pl([35,40])}{Pl([30,40])} = 1 - \frac{\sum_{C\cap(B\cap A)\neq\emptyset} m(C)}{1} = 0.5. \end{aligned}$$

It is clear that the belief–plausibility intervals obtained by this conditioning rule are contained within those obtained by (11)–(12). In fact, Dempster conditioning operates a revision of the disjunctive random set, while the previous conditioning only computes a prediction within a given context. The two conditioning methods coincide when for all focal sets $E$, either $E \subseteq A$ or $E \subseteq A^c$, which is the case in the application of the generalized Bayes theorem [26].

**Remark 4.1.** Interestingly the conditioning rule for conjunctive random sets comes down to the previous conditioning rule (9) 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. Independence

The proper notion of independence between two random sets also depends on the specific interpretation (conjunctive or disjunctive) considered.

*Random set independence* Under the conjunctive approach, two random sets are assumed to be independent if they satisfy the classical notion of independence, when considered as mappings whose "values" are subsets of the final space. Mathematically, two random sets taking values respectively in the power sets of the finite universes[4] $S_1$ and $S_2$ are said to be independent if they satisfy the equalities:

$$P(X = A, Y = B) = P(X = A) \cdot P(Y = B), \quad \forall A \subseteq S_1,\ B \subseteq S_2. \tag{14}$$

Under the above condition, some noticeable probabilities can be also factorized, i.e.:

$$\begin{aligned} P\big(\{X\times Y \subseteq A\times B\}\big) &= \sum_{C\times D\subseteq A\times B} P(X = C)\cdot P(Y = D) \\ &= P(X\subseteq A)\cdot P(Y\subseteq B), \quad \forall A, B, \end{aligned} \tag{15}$$

[4] The definition of independence between two random sets can be easily extended to the general case where either the initial space $\Omega$, or the final universes, $S_1$ and $S_2$, are necessarily finite. We just need to consider appropriate $\sigma$-fields, $\sigma \subseteq \wp(\Omega)$, $\sigma_1 \subseteq \wp(\wp(S_1))$ and $\sigma_2 \subseteq \wp(\wp(S_2))$ and assume that the two-dimensional vector $(X, Y): \Omega \to \wp(S_1)\times\wp(S_2)$ (whose images are pairs of subsets of $S_1$ and $S_2$, respectively) is $\sigma - \sigma_1 \otimes \sigma_2$ measurable, so that it induces a probability measure on the product $\sigma$-field. In our particular case, we consider by default the three discrete $\sigma$-fields $\wp(\Omega)$, $\wp(\wp(S_1))$ and $\wp(\wp(S_2))$, respectively. For the sake of simplicity, we will not get into further details about the general definition.

$$P\big(\{X \times Y \cap A \times B \neq \emptyset\}\big) = \sum_{C \times D \cap A \times B \neq \emptyset} P(X = C) \cdot P(X = D)$$
$$= P(X \cap A \neq \emptyset) \cdot P(Y \cap B \neq \emptyset), \quad \forall A, B. \tag{16}$$

They are random set counterparts of belief and plausibility functions.

**Example 5.** Let us consider two random sets defined on a population of people. Random set $X$ represents the set of spoken languages (one of them being $e =$ English), and it assigns, to each randomly selected person, the set of languages (s)he can speak. Random set $Y$ assigns to each person the set of countries (s)he has visited during more than two months (we will respectively denote by $s$ and $u$ the USA and the UK). Let us take a person at random and consider the events "(S)he speaks (at least) English" and "(S)he has visited the USA or the UK for more than two months". Those events can be mathematically expressed as $X \ni e$ and $Y \cap \{s, u\} \neq \emptyset$, respectively. If they are not independent, i.e., if the probability of the conjunction $P(X \ni e, Y \cap \{s, u\} \neq \emptyset)$ does not coincide with the product of the probabilities of both events, $P(X \ni e)$ and $P(Y \cap \{s, u\} \neq \emptyset)$, then we claim that $X$ and $Y$ are stochastically dependent. The above probabilities can be written in the language of plausibility measures as $Pl(\{e\} \times \{s, u\})$, $Pl_1(\{e\})$ and $Pl_2(\{s, u\})$ respectively (although this is not the correct interpretation here). These three values do not satisfy Eq. (16), and therefore it is clear that the random sets $X$ and $Y$ are not stochastically independent.

Let the reader notice that, when two random sets are, in fact, random intervals, they are independent if and only if the random vectors determined by their extremes are stochastically independent, as we illustrate in the following example.

**Example 6.** Let the random intervals $X = [x_1, x_2]$ and $Y = [y_1, y_2]$ respectively represent the interval of min–max temperatures and min–max humidity on a day selected at random in some location. $X$ and $Y$ would be independent if and only if the random vectors $(x_1, x_2)$ and $(y_1, y_2)$ were stochastically independent, i.e., if and only if they would satisfy:

$$P(x_1 \leqslant a, x_2 \leqslant b, y_1 \leqslant c, y_2 \leqslant d) = P(x_1 \leqslant a, x_2 \leqslant b) \cdot P(y_1 \leqslant c, y_2 \leqslant d), \quad \forall a, b, c, d \in \mathbb{R}. \tag{17}$$

*Independence between ill-known variables* Suppose that we observe a pair of random variables $x$ and $y$ with imprecision, and our information about each realization $(x(\omega), y(\omega))$ is represented by a pair of sets, $(X(\omega), Y(\omega))$, containing the actual pair of values. Given an arbitrary pair of events $A \subseteq S_1$ and $B \subseteq S_2$, the probability values $p_X(A)$ and $p_Y(B)$ are known to belong to the respective intervals of values: $[\underline{P}_X(A), \overline{P}_X(A)]$ and $[\underline{P}_Y(B), \overline{P}_Y(B)]$, where $X$ and $Y$ denote the multivalued mappings that represent our incomplete information about $x$ and $y$, respectively, and $\underline{P}_X, \overline{P}_Y, \underline{P}_Y, \overline{P}_Y$ denote their respective lower and upper probabilities, in the sense of Dempster. If we know, in addition, that both random variables are independent, then we can represent our knowledge about the joint probability $p_{(x,y)}(A \times B)$ by means of the set of values:

$$\big\{p \cdot q\colon\ p \in \big[\underline{P}_X(A), \overline{P}_X(A)\big] \text{ and } q \in \big[\underline{P}_Y(B), \overline{P}_Y(B)\big]\big\}.$$

According to [13], the family of joint probability measures:

$$\mathcal{P} = \big\{P\colon\ P(A \times B) = p \cdot q \text{ s.t. } p \in \big[\underline{P}_X(A), \overline{P}_X(A)\big], q \in \big[\underline{P}_Y(B), \overline{P}_Y(B)\big],\ \forall A, B\big\}$$

is included in the family of probabilities dominated by the joint plausibility $Pl$ considered in Eq. (16). Furthermore, according to [11] it is, in general, a proper subset of it. This fact does not influence the calculation of the lower and upper bounds for $P_{(x,y)}(A \times B)$, but it impacts further calculations about bounds of other parameters associated to the probability measure $P_{(x,y)}$ such as the variance or the entropy of some functions of the random vector $(x, y)$, as it is checked in [14].

As illustrated above, random set independence is the usual probabilistic notion in the conjunctive setting. But it has also a meaningful interpretation within the disjunctive framework: in this case, the pair of set-valued mappings $(X, Y)$ indicates incomplete information about an ill-observed random vector $(x, y)$. Independence between $X$ and $Y$ indicates independence between incomplete pieces of information pertaining to the attributes (the sources providing them are independent), something that has nothing to do, in general, with the stochastic relation between the actual attributes $x$ and $y$ themselves. We will illustrate this issue with some examples.

**Example 7.** Let us suppose that the set-valued mapping $X$ expresses vacuous information about an attribute $x$. In such a case, $X$ will be the constant set-valued mapping that assigns the whole universe, to any element $\omega$ in the initial space, and therefore, it is independent from any other set-valued mapping. But this fact does not mean at all that the attribute $x$ is independent from every other attribute $y$. Let us consider for instance the users of a crowded airport, and let $x$ and $y$ respectively denote their height, in cm, and their nationality. Let us suppose that, on the one hand, we have precise information about their nationality, but on the other hand, we have no information at all about their height. We can represent our information about their height by means of the constant set-valued mapping $X$ that assigns the interval $[90, 250]$ to any passenger. On the other hand, our information about the nationality can be represented by means of a function whose images are singletons ($Y(\omega) = \{y(\omega)\},\ \forall \omega \in \Omega$). Both set-valued mappings satisfy the condition of random set independence but this does not mean that height and nationality are two independent attributes in the considered population.

The next example shows the converse situation: now, the ill-known information about a pair of independent random variables is represented by means of a pair of stochastically dependent random sets.

**Example 8.** Let us consider again the situation described in Example 3. We can easily check that the random sets $X_1$ and $X_2$ are stochastically dependent, according to their joint probability distribution provided in Table 1. Let us now suppose that somebody else considers the following procedure in order to paint the unpainted balls: once both balls have been selected, we drop two coins to decide their color (if they are unpainted), one coin for each ball. If we respectively denote by $x_1$ and $x_2$ the final colors of both ball, we can easily check that $p_{(x_1,x_2)}(\{(r,r)\}) = p_{(x_1,x_2)}(\{(r,w)\}) = p_{(x_1,x_2)}(\{(w,r)\}) = p_{(x_1,x_2)}(\{(w,w)\}) = 0.25$, so the joint probability induced by $(x_1, x_2)$ can be factorized as the product of its marginals.

In the above example, both attributes (the actual colors of both randomly selected balls) are independent random variables, while the initial information about them is represented by means of two stochastically dependent random sets. As we deduce from Examples 7 and 8, neither independence between pieces of (incomplete) information about two attributes implies independence between them, nor the converse is true.

The next example goes along the same line as the last one, and it illustrates a situation a little bit closer to reality.

**Example 9.** Suppose that we have a light sensor that displays numbers between 0 and 255. We take 10 measurements per second. When the brightness is higher than a threshold (255), the sensor displays the value 255 during 3/10 seconds, regardless the actual brightness value. Below we provide data for six measurements:

| Actual values | 215 | 150 | 200 | 300 | 210 | 280 |
|---|---|---|---|---|---|---|
| Displayed quantities | 215 | 150 | 200 | 255 | 255 | 255 |
| Set-valued information | $\{215\}$ | $\{150\}$ | $\{200\}$ | $[255,\infty)$ | $[0,\infty)$ | $[0,\infty)$ |

The actual values of brightness represent a realization of a simple random sample of size $n = 6$, i.e., of a vector of six independent identically distributed random variables. Notwithstanding, our incomplete information about them does not satisfy the condition of random set independence. In fact, we have:

$$P\big(X_i \supseteq [255,\infty) | X_{i-1} \supseteq [255,\infty), X_{i-2} \not\supseteq [255,\infty)\big) = 1, \quad \forall i \geqslant 3.$$

In summary, when several random set-valued mappings indicate incomplete information about a sequence of random (point-valued) measurements of an attribute, independence between them would indicate independence between the sources of information, that we should distinguish from the actual independence between the actual outcomes. The difference between the idea of independence of several random variables and independence of the random sets used to represent incomplete information about them impacts the studies of Statistical Inference with imprecise data: according to the above examples, a sequence of random sets indicating imprecise information about a sequence of i.i.d. random variables does not necessarily satisfy the property of random set independence between its components. In fact, the family of product probabilities, each of which is dominated by an upper probability is strictly contained [11] in the family of probabilities dominated by the product of these upper probabilities (plausibilities, when the universe is finite). In other words, considering a sequence of i.i.d. random sets would not be appropriate when we aim to represent incomplete information about a sequence of i.i.d. random variables.

*Source independence in evidence theory* A similar epistemic independence notion is commonly assumed in the theory of belief functions, when the focal sets of a joint mass assignment $m$ (the mass assignment associated to the pair of ill-known quantities $x$ and $y$ ranging on $S_1$ and $S_2$, respectively) are Cartesian products of subsets of $S_1$ and $S_2$ and that, furthermore, $m$ can be factorized as the products of its marginals:

$$m(C \times D) = m_1(C) \cdot m_2(D), \quad \forall C \subseteq S_1,\ D \subseteq S_2, \tag{18}$$

which is formally the same as Eq. (15). Under the above condition, both the plausibility and the belief function associated to $m$ can be also factorized as the product of their marginals, just rewriting equalities (15)–(16)

$$Bel(A \times B) = Bel_1(A) \cdot Bel_2(B), \tag{19}$$

$$Pl(A \times B) = Pl_1(A) \cdot Pl_2(B). \tag{20}$$

In fact this notion of independence is used to justify the conjunctive rule of combination of Smets, in the case where $x = y$, which is an extreme dependence situation between variables, and masses $m_1$ and $m_2$ are pieces of information about the same quantity $x$ coming from epistemically independent sources. Then $m(C \times D)$ is assigned to $C \cap D$, thus building the mass function $m_\cap(E) = \sum_{C\cap D=E} m(C \times D)$. Conditioning the result on the assumption that the mass assigned to empty set should be zero yields Dempster rule of combination. In the language of random sets, if $X_1$ and $X_2$ are random sets describing the pieces of information forwarded by sources 1 and 2 about $x$, it would read

$$P(X_1 \cap X_2 = E \mid X_1 \cap X_2 \neq \emptyset) = \frac{P(X_1 \cap X_2 = E)}{P(X_1 \cap X_2 \neq \emptyset)} = \frac{\sum_{C \cap D = E} p_{X_1}(C) p_{X_2}(D)}{\sum_{C \cap D \neq \emptyset} p_{X_1}(C) p_{X_2}(D)}.$$

## 5. When the meaning of the model affects the result: interval data processing

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 [10] or stating a regression problem.

### 5.1. Empirical variance for random interval data

Consider a data set consisting of a bunch of intervals $\mathbb{D} = \{I_i = [\underline{a}_i, \bar{a}_i], i = 1, \ldots, n\}$. When computing an empirical variance, 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?

*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:

$$ScalVar(\mathbb{D}) = \frac{\sum_{i=1,\ldots,n} d^2(M, I_i)}{n}, \tag{21}$$

where $M = [\sum_{i=1}^{n} \underline{a}_i/n, \sum_{i=1}^{n} \bar{a}_i/n]$ is the interval mean value, and $d$ is a scalar distance between intervals (e.g. 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. This is the variance suggested by Fréchet for random elements defined on a complete metric space. In fact, it can be extended to random (fuzzy) sets of any dimension by considering, for instance, the families of distances introduced in [41] or [71].

*Epistemic interval data* Under the epistemic view, each interval $I_i$ stands for an ill-known precise value $y_i$. It can be the result of a repeated measurement process subject to randomness and imprecision (each measurement is imprecise). Then we may be 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

$$EVar_1(\mathbb{D}) = \{var(\{y_i, i = 1, \ldots, n\}) \colon\ y_i \in I_i,\ \forall i\}. \tag{22}$$

Computing this interval is a non-trivial task [36,46,64].

*Epistemic interval random data* Alternatively one may consider that the quantity $y$ that we wish to describe is intrinsically random. Each measurement process is an information source providing incomplete information on the variability of $y$. 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

$$EVar_2(\mathbb{D}) = \left\{ var\left(\sum_{i=1}^{n} P_i/n\right) \colon\ SUPP(P_i) \subseteq I_i,\ \forall i = 1, \ldots, n \right\} \tag{23}$$

and it is easy to see that $EVar_1(\mathbb{D}) \subset EVar_2(\mathbb{D})$. In the extreme case of observing a single epistemic interval $(y, [a, b])$, if $y$ is a deterministic ill-known quantity, it has a unique true value. Then $EVar_1([a, b]) = var(y) = 0$ (since even if ill-known, $y$ is not supposed to vary: the set of variances of a bunch of Dirac functions is $\{0\}$). In the second case, $y$ is tainted with variability, $var(y)$ is ill-known and lies in the interval $EVar_2([a, b]) = [0, v^*]$ where $v^* = \sup\{var(y), SUPP(P_y) \subseteq [a, b]\} = (b - a)^2$. 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 nevertheless 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 (as measured by their respective lengths), or by an estimate of the actual variance of the underlying quantity, by computing the variance of say the mid-points of the intervals. But note that as shown in [10], the variance of the mid-points is not necessarily a lower bound, nor an upper bound for the actual value of the variance. Recently suggested scalar variances [60] between intervals come down a linear combination of such a scalar variability estimation and the variance of imprecision.

### 5.2. Regression for interval data

The recent literature contains different generalizations of classical statistical techniques such as point and interval estimation, hypothesis testing, regression, classification, etc., where point-valued attributes are replaced by set-valued ones. Depending on the interpretation of those intervals, some proposals are more appropriate than others. Bearing in mind the difference between the disjunctive and conjunctive reading of sets, two completely different approaches that generalize hypothesis testing techniques have been already reviewed in [16]. Regarding other techniques such as parameter estimation or classification, the disjunctive-approach literature is much wider than the conjunctive one. However, the literature on the regression problem is probably more balanced, and we can find many contributions from both sides. We will therefore use this particular problem as representative of the different statistical techniques in order to illustrate the difference between the disjunctive and the conjunctive approaches.

In the classical least squares method, we consider a collection of vectors of the form $(x_i, y_i)$ and we try to find the function $f : \mathbb{R} \to \mathbb{R}$ that minimizes the sum of the squares of the errors $\sum_{i=1}^{n}(y_i - f(x_i))^2$. Here, the explained variable will be assumed to be interval-valued, while the explanatory variable will take values in the real line. Let us therefore consider a set of bidimensional numerical interval data $\mathbb{D} = \{(x_i, Y_i = [\underline{y}_i, \overline{y}_i]), i = 1, \ldots, n\}$ or its fuzzy counterpart (if the $Y_i$'s become fuzzy intervals). 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.

*Fuzzy least squares* A first approach that is widely known is Diamond's fuzzy least squares method [22]. 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 $\sum_{i=1}^{n} d^2(Ax_i + B, Y_i)$, 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 a complete metric space. 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) interval data as ontic entities.
- If the (fuzzy) interval data set is epistemic, the use of $\tilde{Y} = \tilde{A}^*x + \tilde{B}^*$ presupposes that not only the relation between $x$ and $y$ is linear, but also the evolution of the informativeness of knowledge is linear with $x$.
- 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 various scalar distances exist (see [35] for a recent one) including extensions to fuzzy-valued inputs [38]. These approaches all adopt the ontic view.

*Possibilistic regression* Another classical approach was proposed by Tanaka et al. in the early 1980's (see [70] 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 $Y_i \subset Y(x_i), i = 1, \ldots, n$. Some comments help situate this method:

- It does not presuppose a specific (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 the amount 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 has an affine behavior does not imply that the knowledge about it in each point should be also of the form $Y(x) = Ax + B$. This is similar with the difference between the independence between variables and the independence between sources of information. In an ontic reading, one may wish to interpolate the interval data in a more sophisticated way than assuming linearity of the boundaries of the evolving interval (see Boukezzoula et al. [4] for improvements of Tanaka's methods that cope with such defects).

*Quantile regression* 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 [58,67]. 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 [27]. Varying $\alpha$ 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 $x_i$.

*Sensitivity analysis* A very different approach consists in performing a sensitivity analysis yielding all regression results one would obtain from all precise datasets $\vec{d}$ consistent with $\mathbb{D}$. The aim is to find the range of results one would have obtained with linear regression, had the data been precise. Strangely enough, this technique is seldom considered. Formally it can be posed as follows: Find

$$\begin{aligned} Y(x) &= \{\hat{a}(\vec{d})x + \hat{b}(\vec{d}) \colon \vec{d} \in \mathbb{D}\} \\ &= \left\{\hat{a}x + \hat{b},\ \forall \hat{a}, \hat{b} \text{ that minimize } \sum_{i=1}^{n}(ax_i + b - y_i)^2,\ \forall y_i \in [\underline{y}_i, \overline{y}_i],\ i = 1, \ldots, n\right\}. \end{aligned}$$

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 [51,52].

Yet another approach that can be also framed into the epistemic view has been considered in [65,66,68]. There, an interval-valued mean quadratic error (the set of all possible values for the mean quadratic error) is considered, and different algorithms in order to "minimize" it are built under different conditions, according to some criterion of preference between intervals. The numerical computations seem to be more feasible than those in the original problem (ranging all the possible sample datasets, computing all the regression lines compatible with them, and finally computing the bounds of the response variable according to them).

## 6. 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 [31], a membership function is an abstract object that needs to be interpreted in practical settings in order to be used appropriately. 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 [76] 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. In this section we focus on the difference between ontic and epistemic fuzzy sets. In the latter case, membership grades evaluate degrees of plausibility and the underlying quantity can again be a constant or a variable.

### *6.1. Ontic vs. epistemic fuzzy sets*

As already acknowledged a long time ago, fuzzy sets, like sets, may have a conjunctive or a disjunctive reading [29,78,74]. 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 measurement 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, fuzzy regions in an image, fuzzy rating profiles according to various attributes. As a concrete example, consider the fuzzy set of languages more or less well spoken by a person; then degrees of membership can be interpreted as degrees of proficiency in languages.

In contrast, Zadeh [77] also proposed to interpret membership functions as possibility distributions, paving the way to a representation of incomplete information along a line pioneered thirty years earlier by Shackle [61]. In that case, a degree of membership refers to the idea of plausibility in the sense of Shafer. A possibility distribution, denoted by $\pi$, 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 $\pi$ induces a possibility measure $\Pi$ on $2^S$ such that $\Pi(A) = \sup_{s \in A} \pi(s)$ for all events $A$ and a necessity measure $N(A) = 1 - \Pi(A^c)$ [28]. Function $N$ is also a special (consonant) case of belief function for which all focal sets are nested and the useful information to reconstruct it is contained in the contour function $\pi(s) = Pl(\{s\})$.

### 6.2. What epistemic fuzzy sets can express

Now, if the information about a quantity $x$ is expressed by means of a fuzzy set, the distinction between the deterministic and the stochastic case, discussed for set-representations, is again at work. If $x$ is deterministic, then fuzzy sets must be interpreted in terms of "confidence sets" [12,27] as follows. Let $E_\alpha = \{s, \pi(x) \geqslant \alpha\}$ be the $\alpha$-cut of $\pi$:

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 [73], $1-\alpha$ is the maximal price at which this expert would buy the gamble that wins $1 if the real value of $x$ actually lies in $E_\alpha$ (the minimal selling price for this gamble is $1). Note that there is no "real probability distribution" underlying $\pi$, but Dirac functions as $x$ is deterministic. The consonance of the family of sets $E_\alpha$ 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 $\pi$.

- Mathematically speaking, a possibility measure is a coherent upper probability [73], namely $\Pi(A) = \sup_{P \in \mathcal{P}_\pi} P(A)$ where $\mathcal{P}_\pi = \{P, \forall A, P(A) \leqslant \Pi(A)\}$. So, $\pi$ encodes the set of probabilities $\mathcal{P}_\pi$ [30,24]. 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 $\pi$ understood as $\mathcal{P}_\pi$ 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 $\pi$ as encoding a higher-order (subjective) possibility distribution restricting a family of higher-order subjective probabilities $P_s$ over a set of objective probabilities $P_x$ describing the behavior of a random variable $x$. Namely, $\pi$ is induced by all constraints of the form $P_s(P_x(E_\alpha) = 1) \geqslant 1-\alpha, \forall\alpha > 0$, in other words:

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

  So the domain of $\pi$ can be canonically extended to the set of probability measures on $S$ as follows: $\pi(P_x) = \sup\{\alpha, P_x \text{ has support in } E_\alpha\}$. The possibility measure $\Pi$ is a "second-order possibility" formally equivalent to those considered in [18]. This terminology is used because it is a possibility distribution defined over a set of probability measures. It would be interesting to investigate the relationship between the set of probabilities $\mathcal{P}_\pi$ and the higher-order possibility model. Such a set of probabilities can be equivalently represented by means of a second-order Boolean possibility measure (a 0–1-valued possibility measure defined over the set of probability measures).

### 6.3. Examples of epistemic fuzzy sets

In summary we can consider crisp or fuzzy epistemic sets describing our knowledge of a quantity that can be deterministic or random. According to this rationale, we can distinguish four different situations, and we can use several equivalent representations in each of them. Below, we provide a short formal account of each situation, illustrated with simple examples.

**1A** A crisp set $E$ models our incomplete knowledge of an otherwise fixed quantity $x$. All we know about $x$ is that it belongs to $E$.[5]

| Representation | Deterministic | Random |
|---|---|---|
| Set | $\boldsymbol{E}$ | $\delta_E = \{\delta_x \colon x \in E\}$ |
| Possib. distrib. | $\pi(x) = \begin{cases} 1, & \text{if } x \in E \\ 0, & \text{if } x \notin E \end{cases}$ | $\pi(P) = \begin{cases} 1, & \text{if } P \in \delta_E \\ 0, & \text{if } P \notin \delta_E \end{cases}$ |

**Example 10.** We measure the weight of an object on a scale, whose precision is within a 10 g error margin. If we observe the displayed quantity $d_0$, our information about the actual weight, $x$, can be described by means of the interval $E = [d_0 - 10, d_0 + 10]$.

**1B** A crisp set $E$ models our incomplete knowledge about a random variable $x$. All we know about the probability induced by $x$ is that the support is included in $E$.

[5] The simplest representation here and in the next three cases will be indicated in bold text.

| Representation | Random only |
| --- | --- |
| Set | $\boldsymbol{\mathcal{P}_E = \{P\colon\ P(E) = 1\}}$ |
| Possib. distrib. | $\pi(P) = \begin{cases} 1, & \text{if } P \in \mathcal{P}_E \\ 0, & \text{if } P \notin \mathcal{P}_E \end{cases}$ |

**Example 11.** Let us consider the scale of Example 10. We assume that the observed value on the scale, say $d$, is noisy for a fixed object. If we choose the same object again, our measurement could change. These differences are attached to the randomness of the measurement process. Here, we are interested in describing our knowledge about the probability distribution of the random quantity $d$, based on a single measurement resulting in the outcome $d_0$. This is different from the previous example where the measured value (the estimated weight of an object) is not random, but not totally observable due to the lack of precision of the measurement device. Here our knowledge about the true weight $x$ is affected by both noise and lack of precision. Our information about the probability distribution of $d$ is only given by the set of probability measures whose support is included in $E = [d_0 - 10, d_0 + 10]$, so $\mathcal{P}_E = \{P\colon\ P([d_0 + 10, d_0 + 10]) = 1\}$. But since each of the observations of $x$ is also imprecise with precision like in Example 10, when we observe $x = d_0$, we end up only knowing that $x \in [d_0 - 20, d_0 + 20]$,

**2A** A fuzzy set $\tilde{E}$ models our incomplete knowledge of an otherwise fixed quantity $x$. The membership degree of each element in the universe reflects the degree of possibility that $x$ coincides with it.

| Representation | Deterministic | Random |
| --- | --- | --- |
| Set | $\tilde{E}$ | $\tilde{E}$ over Dirac measures |
| Possib. distrib. | $\boldsymbol{\pi(x) = \sup\{\alpha\colon\ x \in \tilde{E}_\alpha\}}$ | $\pi(\delta_x) = \sup\{\alpha\colon\ x \in \tilde{E}_\alpha\}$ |

**Example 12.** The accuracy expected to be obtained using a GPS receiver will vary according to the overall system used. According to the usual GPS specifications, the information about the current position, $\mathbf{x} = (x_1, x_2)$ can be described by means of a family of nested circles ($\alpha$-cuts), all of them centered at the displayed position $\mathbf{d}$ and each one of them attached to a confidence degree:
- The confidence degree of $E_{0.5}$ is 0.5, $E_{0.5}$ being called the "circular error probability (CEP)".
- The circle $E_{0.32}$ is associated to the confidence degree 0.68 and it is "root mean square (RMS)".
- The confidence degree of $E_{0.05}$ is 0.95. It is the so-called "95% radius (R95)".
- The circle $E_{0.02}$ is associated to a confidence degree 0.98, and it is named "twice the distance root mean square (2RMS)".

According to [12,27], the above family of confidence restrictions can be alternatively represented by means of a fuzzy set $\pi$. Each value $\pi(\boldsymbol{y})$ denotes the possibility that the actual (fixed) position $\boldsymbol{x}$ coincides with $\boldsymbol{y}$.

**2B** A fuzzy set $\tilde{E}$ models our incomplete knowledge of a random variable $x$. The fuzzy set represents a possibility distribution encoding a family of subjective probabilities. Alternatively, each $\alpha$-cut of the fuzzy interval may be viewed as containing the support of the distribution with (subjective) probability at least $1 - \alpha$.

| Representation | Random only |
| --- | --- |
| Set | $\boldsymbol{\mathcal{P}_\pi = \{P_x\colon\ P_x(A) \leqslant \sup_{a \in A} \pi(a),\ \forall A\}}$ |
| Crisp possib. distrib. | $\pi(P_x) = \begin{cases} 1, & \text{if } P_x \in \mathcal{P}_\pi \\ 0, & \text{if } P_x \notin \mathcal{P}_\pi \end{cases}$ |
| Gradual possib. distrib. | $\pi(P_x) = \sup\{\alpha\colon\ P_x(E_\alpha) = 1\}$ |

**Example 13.** Let us consider the information provided in Example 12 and assume that the displayed point $\boldsymbol{d}$ may vary for a fixed location $\boldsymbol{x}$. These differences are attached to the randomness of the measurement process, that depend on the current position of satellites. Here, we are interested in describing our knowledge about the probability distribution of the displayed location $\boldsymbol{d}$ in a forthcoming measurement, based on a single measurement resulting in the outcome $d_0$. Our information about the probability distribution of $\boldsymbol{d}$ is given by the set of probability measures dominated by the possibility measure $\Pi$ determined by the following equation:

$$\Pi(A) = \sup_{x \in A} \pi(x), \quad \forall A \subseteq \mathbb{R}^2,$$

where $\pi$ is a fuzzy set very closely related to the one considered in Example 12: its $\alpha$-cuts are also centered on the displayed position $\boldsymbol{d}$ and their diameters are twice the diameters of the cuts of the fuzzy set $E$.

The simple use of a fuzzy set interpreted as a possibility distribution dominating an objective probability distribution does not make it clear where the objective probability distribution comes from, i.e. the underlying sample space. Moreover, it does not account for a possible underlying measurement process of $x$. Namely, regardless of whether $x$ is deterministic or stochastic, there may be a stochastic measurement process yielding more or less accurate 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.

### 6.4. 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 [48,49] in the late seventies, clearly underlying an epistemic view of fuzzy sets, a line followed up by Kruse and Meyer [47]. 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 $(\Omega, P, X)$ 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

$$\pi(x) = \inf_{\omega \in \Omega} \mu_{\tilde{X}(\omega)}\big(x(\omega)\big). \tag{24}$$

For each level $\alpha \in (0, 1]$, $\tilde{X}_\alpha(\omega) = \{s \in S: \ \mu_{\tilde{X}(\omega)}(s) \geqslant \alpha\}$ is a multiple-valued mapping such that $(\Omega, P, \tilde{X}_\alpha)$ is an epistemic random set according to Dempster framework. Kruse and Meyer [47] 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 [1,14]. The evidence theory counterpart of this view deals with belief functions having fuzzy focal elements [75]. An alternative epistemic view of fuzzy random variables was more recently proposed in the spirit of Walley [73], in terms of a convex set of probabilities induced on $S$ [15].

In contrast, the line initiated in the mid-1980's by Puri and Ralescu [59] 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 [7,39]. The main issue is to define a space of functions equipped with a suitable metric structure [23,71]. 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 [41], 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. The existence of two views of fuzzy random variables, the one initiated by Kwakernaak and the one proposed by Puri and Ralescu, is acknowledged, surveyed and discussed in [37]. See [10] for an extensive comparison of the three views of fuzzy random variables, including the one based on imprecise probabilities.

The ontic view is advocated by Colubi et al. [8] in the statistical analysis of linguistic data. The authors argue that they are interested in the statistics of perceptions (see [3] for a general presentation and defense of this paradigm). 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 considered experiment involves a simultaneous double rating: a free-response fuzzy rating along with a linguistic one. The alleged goal is to predict the category that a person considers correct for the segment. The linguistic evaluation is performed to validate the classification process introduced in the paper. The precise length of the segment exists but it is considered by these authors to be irrelevant for the classification goal. These authors concede that to predict the real length from the fuzzy perceptions requires a different approach. While the linguistic labels can arguably be considered as ontic entities, the fuzzy rating with free format can hardly be so, as it really is a numerical rendering of the imprecision pervading the perception of an actual length.

The case of Likert scaling is more problematic. This is a method of ascribing quantitative values to qualitative data, thus making 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 values 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 [25]. 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 formalize fuzzy random set methods, even if there exist programs in the language R which make them easier to apply [72]. It may be that ordinal statistical methods devoted to finite qualitative scales would be more appropriate in this case.

A major application of fuzzy random variables is regression analysis. This topic would deserve an extensive discussion but this is beyond the scope of this paper. In any case such a discussion would follow the same lines as the one above concerning interval regression. In fact, most current fuzzy set approaches are of the ontic style, following the pioneering work of Diamond [22], or adopt the possibilistic view of Tanaka [70]. Recently, Denoeux's [20] generalized EM algorithm explicitly adopts an epistemic view of fuzzy data. But his approach, based on Zadeh's probability of a fuzzy event, yields a precise regression line, based on averaging. This is in agreement with the philosophy of the EM algorithm and it contrasts with the idea of propagating the imprecision modelled by fuzzy data on the results, as done in [52] as well as with the idea of representing the imprecision of precise noisy data as done in [58].

## 7. 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 versions of old 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 optimize 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. For instance, the study of fuzzy differential inclusions in the style initiated by Huellermeier [44] goes in this direction (see [6]).

## Acknowledgements

This work has been partially supported by the Spanish Ministry of Education and Science and the European Regional Development Fund (FEDER), under project TIN2011-24302.

## References

[1] 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.
[2] A.S.C. Bertoluzza, A. Salas, N. Corral, On a new class of distances between fuzzy numbers, Mathw. Soft Comput. 2 (1995) 71–84.
[3] A. Blanco-Fernández, et al., A distance-based statistical analysis of fuzzy number-valued data, Int. J. Approx. Reason. 55 (7) (2014) 1487–1501.
[4] R. Boukezzoula, S. Galichet, A. Bisserier, A midpoint-radius approach to regression with interval data, Int. J. Approx. Reason. 52 (2011) 1257–1271.
[5] D. Bouyssou, D. Dubois, M. Pirlot, H. Prade (Eds.), Decision-Making Process—Concepts and Methods, ISTE/Wiley, London/New York, 2009.
[6] Y. Chalco-Cano, H. Román-Flores, Some remarks on fuzzy differential equations via differential inclusions, Fuzzy Sets Syst. (2013), http://dx.doi.org/10.1016/j.fss.2013.04.017, in press.
[7] A. Colubi, Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data, Fuzzy Sets and Systems 160 (3) (2009) 344–356.
[8] A. Colubi, G. González-Rodríguez, M.A. Gil, W. Trutschnig, Nonparametric criteria for supervised classification of fuzzy data, Int. J. Approx. Reason. 52 (2011) 1272–1282.
[9] O. Cordón, A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems, Int. J. Approx. Reason. 52 (6) (2011) 894–913.
[10] I. Couso, D. Dubois, On the variability of the concept of variance for fuzzy random variables, IEEE Trans. Fuzzy Syst. 17 (2009) 1070–1080.
[11] I. Couso, S. Moral, Independence concepts in evidence theory, Int. J. Approx. Reason. 51 (2010) 748–758.
[12] I. Couso, S. Montes, P. Gil, The necessity of the strong $\alpha$-cuts, Int. J. Uncertain. Fuzziness Knowledge-Based Systems 9 (2001) 249.
[13] I. Couso, S. Moral, P. Walley, A survey of concepts of independence for imprecise probabilities, Risk Decis. Policy 5 (2000) 165–181.
[14] I. Couso, L. Sánchez, Higher order models for fuzzy random variables, Fuzzy Sets and Systems 159 (2008) 237–258.
[15] I. Couso, L. Sánchez, Upper and lower probabilities induced by a fuzzy random variable, Fuzzy Sets and Systems 165 (2011) 1–23.
[16] I. Couso, L. Sánchez, Mark-recapture techniques in statistical tests for imprecise data, Int. J. Approx. Reason. 52 (2011) 240–260.
[17] L.M. De Campos, M.T. Lamata, S. Moral, The concept of conditional fuzzy measure, Int. J. Intell. Syst. 5 (1990) 237–246.
[18] G. De Cooman, P. Walley, An imprecise hierarchical model for behaviour under uncertainty, Theory and Decision 52 (2002) 327–374.
[19] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[20] T. Denoeux, Maximum likelihood estimation from fuzzy data using the EM algorithm, Fuzzy Sets and Systems 183 (1) (2011) 72–91.
[21] T. Denoeux, Z. Younes, F. Abdallah, Representing uncertainty on set-valued variables using belief functions, Artificial Intelligence 174 (7–8) (2010) 479–499.
[22] P. Diamond, Fuzzy least squares, Inform. Sci. 46 (1988) 141–157.
[23] P. Diamond, P. Kloeden, Metric Spaces of Fuzzy Sets, World Scientific, Singapore, 1994.
[24] D. Dubois, Possibility theory and statistical reasoning, Comput. Statist. Data Anal. 51 (2006) 47–69.
[25] D. Dubois, The role of fuzzy sets in decision sciences: Old techniques and new directions, Fuzzy Sets and Systems 184 (2011) 3–28.
[26] D. Dubois, T. Denoeux, Conditioning in Dempster–Shafer theory: Prediction vs. revision, in: T. Denoeux, M.-H. Masson (Eds.), Belief Functions, in: Advances in Soft Computing, vol. 164, Springer, 2012, pp. 385–392.
[27] D. Dubois, L. Foulloy, G. Mauris, H. Prade, Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities, Reliab. Comput. 10 (2004) 273–297.
[28] D. Dubois, H. Prade, Possibility Theory, Plenum Press, New York, 1988.
[29] D. Dubois, H. Prade, Incomplete conjunctive information, Comput. Math. Appl. 15 (1988) 797–810.

[30] D. Dubois, H. Prade, When upper probabilities are possibility measures, Fuzzy Sets and Systems 49 (1992) 65–74.
[31] D. Dubois, H. Prade, The three semantics of fuzzy sets, Fuzzy Sets and Systems 90 (1997) 141–150.
[32] D. Dubois, H. Prade, Formal representations of uncertainty, in: [5], 2009, pp. 85–156, Chap. 3.
[33] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets, Fuzzy Sets and Systems 192 (2012) 3–24.
[34] R. Fagin, J.Y. Halpern, A new approach to updating beliefs, in: P.P. Bonissone, M. Henrion, L.N. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, UAI'91, Elsevier Science, New York, 1991, pp. 347–374.
[35] 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.
[36] S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpre, M. Aviles, Computing variance for interval data is NP-hard, SIGACT News 33 (2002) 108–118.
[37] J. Gebhardt, M.A. Gil, R. Kruse, Fuzzy set-theoretic methods in statistics, in: R. Slowinski (Ed.), Fuzzy Sets in Decision Analysis, Operations Research and Statistics, Kluwer Academic Publishers, Boston, 1998, pp. 311–347.
[38] G. González-Rodríguez, A. Blanco, A. Colubi, M.A. Lubiano, Estimation of a simple linear regression model for fuzzy random variables, Fuzzy Sets and Systems 160 (3) (2009) 357–370.
[39] G. González-Rodríguez, A. Colubi, M.A. Gil, Fuzzy data treated as functional data. A one-way ANOVA test approach, Comput. Stat. Data Anal. 56 (4) (2012) 943–955.
[40] J.Y. Jaffray, Bayesian updating and belief functions, IEEE Trans. Syst. Man Cybern. 22 (1992) 1144–1152.
[41] R. Körner, On the variance of fuzzy random variables, Fuzzy Sets and Systems 92 (1997) 83–93.
[42] J.Y. Halpern, R. Fagin, Y. Moses, M.Y. Vardi, Reasoning about Knowledge, MIT Press, Cambridge, 2003 (revised paperback edition).
[43] A. Herzig, J. Lang, P. Marquis, Action representation and partially observable planning using epistemic logic, in: Proc. Int. Joint Conf. on Artificial Intelligence, IJCAI-03, Acapulco, Morgan Kaufmann, San Francisco, 2003, pp. 1067–1072.
[44] E. Huellermeier, An approach to modelling and simulation of uncertain dynamical systems, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 5 (1997) 117–137.
[45] D.G. Kendall, Foundations of a theory of random sets, in: E.F. Harding, D.G. Kendall (Eds.), Stochastic Geometry, Wiley, New York, 1974, pp. 322–376.
[46] V. Kreinovich, G. Xiang, S. Ferson, Computing mean and variance under Dempster–Shafer uncertainty: Towards faster algorithms, Int. J. Approx. Reason. 42 (3) (2006) 212–227.
[47] R. Kruse, K. Meyer, Statistics with Vague Data, D. Reidel, Dordrecht, The Netherlands, 1987.
[48] H. Kwakernaak, Fuzzy random variables – I. Definitions and theorems, Inform. Sci. 15 (1978) 1–29.
[49] H. Kwakernaak, Fuzzy random variables – II. Algorithms and examples for the discrete case, Inform. Sci. 17 (1979) 253–278.
[50] D.V. Lindley, Scoring rules and the inevitability of probability, Int. Stat. Rev. 50 (1982) 1–26.
[51] K. Loquin, D. Dubois, Kriging and epistemic uncertainty: a critical discussion, in: R. Jeansoulin, et al. (Eds.), Methods for Handling Imperfect Spatial Information, Springer, Berlin, Heidelberg, New York, 2010, pp. 269–305.
[52] K. Loquin, D. Dubois, Kriging with ill-known variogram and data, in: A. Hunter (Ed.), Scalable Uncertainty Management, SUM 2010, Toulouse, Deshpande, in: LNAI, vol. 6379, Springer, Berlin, Heidelberg, New York, 2010, pp. 219–235.
[53] G. Matheron, Random Sets and Integral Geometry, J. Wiley & Sons, New York, 1975.
[54] R. Moore, Methods and Applications of Interval Analysis, Studies in Applied Mathematics, SIAM, Philadelphia, 1979.
[55] H.T. Nguyen, On random sets and belief functions, J. Math. Anal. Appl. 65 (1978) 531–542.
[56] J. Paris, The Uncertain Reasoner's Companion, Cambridge University Press, 1994.
[57] F. Pichon, D. Dubois, T. Denoeux, Relevance and truthfulness in information correction and fusion, Int. J. Approx. Reason. 53 (2012) 159–175.
[58] H. Prade, M. Serrurier, Maximum-likelihood principle for possibility distributions viewed as families of probabilities, in: IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2011, Taipei, IEEE, 2011, pp. 2987–2993.
[59] M. Puri, D. Ralescu, Fuzzy random variables, J. Math. Anal. Appl. 114 (1986) 409–422.
[60] A.B. Ramos-Guajardo, M.A. Lubiano, K-sample tests for equality of variances of random fuzzy sets, Comput. Statist. Data Anal. 56 (2012) 956–966.
[61] G.L.S. Shackle, Decision, Order and Time in Human Affairs, 2nd edition, Cambridge University Press, UK, 1961.
[62] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[63] G. Shafer, A. Tversky, Languages and designs for probability, Cogn. Sci. 9 (1985) 309–339.
[64] M. Spadoni, L. Stefanini, Computing the variance of interval and fuzzy data, Fuzzy Sets and Systems 165 (2011) 24–36.
[65] L. Sánchez, I. Couso, J. Casillas, Genetic learning of fuzzy rules based on low quality data, Fuzzy Sets and Systems 160 (17) (2009) 2524–2552.
[66] L. Sánchez, I. Couso, Advocating the use of imprecisely observed data in genetic fuzzy systems, IEEE Trans. Fuzzy Syst. 15 (4) (2007) 551–562.
[67] L. Sánchez, I. Couso, Learning from imprecise examples with GA-P algorithms, Mathw. Soft Comput. 2–3 (5) (1998) 305–319.
[68] L. Sánchez, J. Otero, I. Couso, Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms, Soft Comput. 13 (3) (2009) 307–318.
[69] P. Smets, The normative representation of quantified beliefs by belief functions, Artificial Intelligence 92 (1997) 229–242.
[70] H. Tanaka, P. Guo, Possibilistic Data Analysis for Operations Research, Physica-Verlag, Heidelberg, 1999.
[71] W. Trutschnig, G. González-Rodríguez, A. Colubi, M.A. Gil, A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread, Inform. Sci. 179 (2009) 3964–3972.
[72] W. Trutschnig, M.A. Lubiano, J. Lastra, SAFD – An R package for statistical analysis of fuzzy data in towards advanced data analysis by combining soft computing and statistics, in: C. Borgelt, et al. (Eds.), Studies in Fuzziness and Soft Computing, vol. 285, Springer, 2013, pp. 107–118.
[73] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.
[74] R.R. Yager, Set-based representations of conjunctive and disjunctive knowledge, Inform. Sci. 41 (1987) 1–22.
[75] J. Yen, Generalizing the Dempster–Shafer theory to fuzzy sets, IEEE Trans. Syst. Man Cybern. 20 (1990) 559–569.
[76] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part I, Inform. Sci. 8 (1975) 199–249.
[77] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 1–28.
[78] L.A. Zadeh, PRUF — a meaning representation language for natural languages, Int. J. Man-Mach. Stud. 10 (1978) 395–460.