

On the variability of the concept of variance for fuzzy random variables

Inés Couso

Departamento de Estadística e I.O. y D.M.

Universidad de Oviedo

e-mail: couso@uniovi.es

Didier Dubois

Institut de Recherche en Informatique de Toulouse

Université Paul Sabatier

e-mail: Didier.Dubois@irit.fr

Abstract

Fuzzy random variables possess several interpretations. Historically, they were proposed either as a tool for handling linguistic label information in statistics or to represent uncertainty about classical random variables. Accordingly, there are two different approaches to the definition of the variance of a fuzzy random variable. In the first one, the variance of the fuzzy random variable is defined as a crisp number, that makes it easier to handle in further processing. In the second case, the variance is defined as a fuzzy interval, offering a gradual description of our incomplete knowledge about the variance of an underlying, imprecisely observed, classical random variable. In this work, we first discuss another view of fuzzy random variables, that comes down to a set of random variables induced by a fuzzy relation describing an ill-known conditional probability. This view leads to yet another definition of the variance of a fuzzy random variable, in the context of the theory of imprecise probabilities. The new variance is a real interval, which achieves a compromise between both previous definitions in terms of representation simplicity. Our main objective is to demonstrate, with the help of simple examples, the practical significance of these definitions of variance induced by various existing views of fuzzy random variables.

Keywords: Imprecise probabilities, Fuzzy random variable, Random set, Variance, Second-order possibility measure.

1 Introduction

The concept of fuzzy random variable, that extends the classical definition of random variable, was introduced by Féron [19] in 1976. Later on and sometimes independently, Kwakernaak [29], Puri and Ralescu [43], Kruse and Meyer [28], Diamond and Kloeden [13], proposed other variants. In [23], Krätschmer surveyed all of these definitions and proposed a unified approach. In all of these papers, a fuzzy random variable is defined as a function that assigns a fuzzy subset to each possible output of a random experiment. The different definitions in the literature disagree on the measurability conditions imposed to this mapping, and in the properties of the output space, but all of them intend to model situations that combine fuzziness and randomness.

Since the introduction of this concept, much effort has been devoted to the generalisation of different probabilistic concepts and classical results to the case when outcomes of a random experiment are represented by fuzzy sets. Generalized definitions of descriptive parameters, useful as information summaries for probability distributions, can be divided into two groups. Some authors view extensions of standard parameters as fuzzy values: the expectation [43], the distribution function in a point [6, 28], the variance [27] and the covariance¹ [34]. For yet other authors, the expectation [31], the variance [18, 22, 32], the covariance [18] or the inequality index [33] are defined as crisp values. These scalar definitions sometimes reflect the idea that the fuzzy outcomes of the random process are viewed as primitive objects (like in [43], while the former definitions consider a fuzzy random variable as the imprecise perception of an otherwise precise random variable. The scalar definitions are also easier to handle in practice.

In spite of the great amount of mathematical results about fuzzy random variables, few publications study the different interpretations that could be given to their various definitions. *Yet there is not a unique intuition behind the various definitions originally proposed in the literature. While Féron [19], Puri and Ralescu [43], Diamond and Kloeden [13] view a fuzzy random variable as the extension of a random set, Kwakernaak [29], Kruse and Meyer [28] consider it models the imprecise perception of an ill-known classical random variable. This divergence of views directly impacts the choice of suitable definitions for extensions of traditional descriptive parameters such as variance.*

This work especially focuses on the meaning of different extensions of the concept of variance to fuzzy random variables. For simplicity we restrict

¹We must remark that the concept of fuzzy random variable not only extends the concept of one-dimensional random variable, but also of random n -dimensional vector.

ourselves to the case where the values of the fuzzy random variable are fuzzy subsets of \mathbb{R} . We review two different definitions of variance, found in the literature, namely, a scalar one, and a fuzzy-interval-valued one. Moreover we propose an additional interval-valued definition, cast in the setting of imprecise probabilities, that relies on a third view of a fuzzy random variable.

We lay bare the three specific interpretations of fuzzy random variables and the corresponding definitions of variance. Guided by simple examples, we observe the advantages and drawbacks of each definition in different contexts. In particular, *the meaning of the variance* depends upon assumptions on the nature of the modelled quantity, according to whether it is deterministic or stochastic.

Interestingly, the study of variance enables a clear distinction to be made between the three views of fuzzy random variables, while expectations cannot properly do it. Indeed, the same fuzzy-valued expectation is obtained in the Puri-Ralescu and the Kruse-Meyer views, even if with different rationales and methods. The third view proposed here yields an interval-valued expectation, that is precisely bounded by the upper and lower expectations of the probability family encoded by the fuzzy-valued expectation obtained in the two other views [5]. When the fuzzy random variable is a random set, all three views yield the same interval-valued expectation. This is not the case with variance as we shall see in this paper. In fact as pointed out by Walley ([47] Appendix G), upper and lower expectations induced by a convex probability family (a credal set) are attained on vertices of the credal set, while upper and lower variances are often attained inside the credal set.

The next sections will be devoted to these three different definitions of the variance, namely the scalar, the fuzzy and the interval-valued ones. In Section 5, we discuss possible connections between the two latter notions.

2 Scalar variance of a fuzzy random variables

Let us recall that the variance of a classical random variable X defined on a probability space (Ω, \mathcal{A}, P) is of the form $\text{Var}(X) = \int_{\Omega} [X - E(X)]^2 dP$, $E(X)$ being the expectation of X . $\text{Var}(X)$ can be expressed as the following Lebesgue integral ² with respect to the probability measure P_X induced by X on the real line:

²Since the variance of a classical random variable is a function of its induced probability distribution, we speak of the variance of such probability distribution, as a slight abuse of language.

$$\text{Var}(P_X) = \int_{\mathbb{R}} \left(x - \int_{\mathbb{R}} y dP_X(y) \right)^2 dP_X(x).$$

Let us consider a metric, d , defined over the class of the fuzzy subsets of \mathbb{R} , $\mathcal{F}(\mathbb{R})$, (oftentimes a subclass, typically the set of compact fuzzy subsets of the real line). Many such metrics are surveyed in [24], and examples of them are discussed later on in this section. A fuzzy random variable $\tilde{X} : \Omega \rightarrow \mathcal{F}(\mathbb{R})$ is an \mathcal{A} - $\beta(d)$ -measurable function (here, $\beta(d)$ represents the Borel σ -algebra induced by d .) The expectation $E(\tilde{X})$ of a fuzzy random variable is a fuzzy subset of the real line which has been given several, often equivalent, definitions in the literature (see [25, 28, 43]). It is basically a kind of integral over Ω of the fuzzy-set-valued function \tilde{X} .

The following definition of variance of a fuzzy random variable is widely found in the literature, and is strikingly similar to the classical notion:

Definition 1. *We call scalar variance of \tilde{X} with respect to distance d , the quantity (when it exists)*

$$\text{SVar}(\tilde{X}) = \int_{\Omega} d(\tilde{X}, E(\tilde{X}))^2 dP. \quad (1)$$

In the following several instances of this definition found in the literature are discussed in more details.

2.1 Some metrics underlying the scalar variance

The various existing definitions of variance of a fuzzy random variable that fit the formulation in Equation 1 satisfy $\text{SVar}(\tilde{X}) = 0$ if and only if all fuzzy outcomes are the same (they have the same membership function). They differ in the used metric and in the definition of expectation.

Körner [22] considers Fréchet's definition of expectation [20] for measurable functions taking values in a metric space. The variance of \tilde{X} is defined as the expectation of the squares of the distances of their values to their Fréchet expectation. It is noticeable that Fréchet defines the expectation of a measurable function Z , with values in a metric space (M, d) as a solution $a = E^{(d)}(Z)$, (not necessarily unique) of the problem $\min_{a \in M} E[d(Z, a)]^2$. Körner's definition is valid when the final space is \mathbb{R}^n , with arbitrary $n \in \mathbb{N}$. In the particular case where \tilde{X} is a classical random vector and the chosen distance is Euclidean, the result of this calculation is the moment of inertia. This way, Körner's procedure generalizes, in the n -dimensional case, a concept that may be useful to measure the dispersion of the membership functions that are values of the fuzzy random variable. However this concept

is not directly related to the concept of variance-covariance matrix. Körner [22] checks that Puri and Ralescu's expectation [43] is the only Fréchet expectation for a certain family of metrics defined over the class of compact and normal fuzzy sets of \mathbb{R} .

Lubiano *et al.* [32] introduce a family of variances defined by (1), the considered expectation being also that of Puri-Ralescu, $E_{PR}(\tilde{X})$, and the class of distances being that defined by Bertoluzza *et al.* in [4]. Bertoluzza *et al.* distances are defined for pairs of convex and compact fuzzy sets of the real line $\mathcal{F}(\mathbb{R})$ (fuzzy intervals [15]) as follows:

$$D_{\vec{\lambda}}(A, B) = \sqrt{\int_{(0,1)} [d_{\vec{\lambda}}(A_\alpha, B_\alpha)]^2 d\alpha},$$

where $A_\alpha = \{\omega \in \Omega : A(\omega) \geq \alpha\}$ and B_α denote respectively the (weak) α -cuts of A and B , $\vec{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in [0, 1]^3$ is a triple of weights such that $\lambda_1 + \lambda_2 + \lambda_3 = 1$, and $d_{\vec{\lambda}}(A_\alpha, B_\alpha)$ denotes a weighted combination of distances between upper and lower bounds and midpoint of the intervals:

$$[\lambda_1(\sup A_\alpha - \sup B_\alpha)^2 + \lambda_2(\text{mid}A_\alpha - \text{mid}B_\alpha)^2 + \lambda_3(\inf A_\alpha - \inf B_\alpha)^2]^{1/2},$$

mid denoting the midpoint of the corresponding interval.

2.2 Random linguistic variables

In practical situations, the values of \tilde{X} are often linguistic labels $\mathbb{L} = \{L_1, \dots, L_k\}$ forming a (fuzzy or not) partition of the real line. Only the coarsening \mathbb{L} of the actual range of an underlying usual random variable is used. Families of scalar variances so defined allow us to quantify the dispersion of the (fuzzy, or set-valued) values of \tilde{X} , regarded as a measurable function from Ω to (\mathbb{L}, d) , adopting a classical point of view. $\text{SVar}(\tilde{X})$ thus evaluates the variation across the possible linguistic labels. Let us illustrate the use of the scalar variance with an example of this kind.

Example 1. *A person has a container of apples and he is asked about their weight. He does not use any scale. He just takes an apple in his hand and chooses between “high”, “medium” and “low”, these terms forming the linguistic scale \mathbb{L} . Each of these labels can be viewed as a suitable fuzzy subset of the weight scale. This information can be described by a fuzzy random variable $\tilde{X} : \Omega \rightarrow \mathbb{L}$, where each $\omega \in \Omega$ represents an apple and $\tilde{X}(\omega)$ represents the label assigned to it (high, medium or large). The probability distribution induced by \tilde{X} determines the proportion of times that this person assigns each label. Let us suppose that he assigns the weight “medium” to all the*

apples, because he cannot appreciate any difference of weight among them. Then the scalar variance takes the value 0. Suppose now than another person (a woman) uses finer notions of “high”, “medium” and “low”, and hence, she can give more accurate descriptions of the weights, so she chooses, for instance, each label one-third of the times. Thus, what the first person calls “medium” covers what the second person calls “high”, “medium” and “low”. Then the scalar variance associated to her assignations is much higher. We can assign different scalar variances to the same container of apples, depending on the accuracy of our answers about the weight. We are not especially interested by the variability of the real weights of apples at this stage, just by the variability of the assignment to classes.

However, the general framework for this kind of fuzzy random variable does not assume the (fuzzy) values of \tilde{X} to form a partition, just that these values are elements of a space equipped with a metric, which is at odds with the linguistic interpretability of the terms in \mathbb{L} . In particular, the scalar distance between any two such fuzzy sets measures the difference in specificity as much as in location. On the contrary, restricting to a fuzzy partition where fuzzy sets only weakly overlap, the distance between terms is an approximation of the distance one could compute if our perception was finer. So the family of scalar variances studied here looks somewhat debatable beyond its use for linguistic random variables.

2.3 Some interpretations of the scalar variance of a fuzzy random variable

The scalar variance is in agreement with the view of a fuzzy random variable as accounting for a standard random phenomenon whose fuzzy outcomes are objects of their own, subject to variability (e.g. linguistic labels). In [22] and [32] we can find some interesting properties of the families of variances defined there, shedding more light on their meaning. Let us consider the formulation of the scalar variance in the particular case when \tilde{X} is a random interval (a function whose values are intervals of the final space and which satisfies a suitable measurability condition.) In this case, the definition of Lubiano et al. is of the form:

$$\text{SVar}(\tilde{X}) = \lambda_1 \text{Var}(X_1) + \lambda_2 \text{Var}\left(\frac{X_1 + X_2}{2}\right) + \lambda_3 \text{Var}(X_2)$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and X_1, X_2 are the random variables defined over Ω as $X_1(\omega) = \inf \tilde{X}(\omega)$ and $X_2(\omega) = \sup \tilde{X}(\omega)$, $\forall \omega \in \Omega$, respectively. (Under the measurability conditions imposed to \tilde{X} by the authors, the functions X_1

and X_2 are \mathcal{A} - $\beta(\mathbb{R})$ measurable.) Thus, we can easily check that it can be alternatively written as follows:

$$\text{SVar}(\tilde{X}) = \pi_1 \text{Var}(X_1) + \pi_2 \text{Cov}(X_1, X_2) + \pi_3 \text{Var}(X_2),$$

where $\pi_1 = \lambda_1 + 0.25\lambda_2$, $\pi_2 = 0.5\lambda_2$, $\pi_3 = \lambda_3 + 0.25\lambda_2$, $\lambda_i \geq 0$, $i = 1, 2, 3$. Furthermore, in this particular situation, the family of distances considered by Körner represents a special case of the family of metrics defined by Bertoluzza et al., and so, Körner’s variance also fulfills the above restrictions. Therefore, in the particular case in which λ_2 is zero (thus $\pi_3 = 1 - \pi_1$), the variance of the random set will be a convex linear combination of the variances of their boundaries³. Additionally, if $\pi_3 = (1 - \sqrt{\pi_1})^2$, the variance of \tilde{X} coincides with the variance of the convex linear combination of X_1 and X_2 given by the expression $\sqrt{\pi_1}X_1 + (1 - \sqrt{\pi_1})X_2$. In other words, in this case, for every element in the sample space, ω , we can choose a representative point, $\alpha X_1(\omega) + (1 - \alpha)X_2(\omega)$ (with $\alpha \in [0, 1]$), of the value of the fuzzy random variable, and then calculate the variance of the resulting classical random variable. The idea of computing the scalar variance using a representative substitute point to each fuzzy observation is used by Baudrit et al. [3], as one piece of information to be extracted from the result of a hybrid propagation of fuzzy and probabilistic information through a mathematical model.

3 Fuzzy variance associated to a second-order imprecise model

The scalar variance does not inform much about the variability of some property of the elements of a population Ω , modelled by (classical) underlying random variable, but known only through imprecise observations, a situation again captured by a fuzzy random variable. *This is Kwakernaak’s interpretation, taken over by Kruse and Meyer [28].* They choose a possibilistic interpretation of fuzzy sets. Each fuzzy set is viewed as modeling incomplete knowledge about an otherwise precise value. These authors then claim that the fuzzy random variable represents imprecise or vague knowledge about a *classical* random variable, $X_0 : \Omega \rightarrow \mathbb{R}$, they refer to as the “original random variable.” Therefore, the membership degree $\tilde{X}(\omega)(x)$ of a real number x to the fuzzy set $\tilde{X}(\omega)$ represents the possibility degree of the assertion “ $X_0(\omega)$ is x ”, i.e., the image of element ω coincides with x . Note that while X_0 is

³The definition given by Feng in [18] and cited in the introduction fits this formulation for $\pi_1 = \pi_3 = 0.5$.

random, $X_0(\omega)$ is deterministic, once ω is fixed. On this ground, these authors define a possibility measure over the set of all random variables, which is associated to the so-called “acceptability function”. The “acceptability degree” of each random variable, $X : \Omega \rightarrow \mathbb{R}$, is computed as:

$$\text{acc}_{\tilde{X}}(X) = \inf_{\omega \in \Omega} \tilde{X}(\omega)(X(\omega)).$$

According to Kruse and Meyer ([28]), $\text{acc}_{\tilde{X}}(X)$ represents the grade of possibility that X is the “true” random variable that models the studied experiment. In other words, it represents the grade of possibility of the assertion

$$X_0(\omega) = X(\omega), \quad \forall \omega \in \Omega.$$

3.1 The case of disjunctive random sets

When, in particular, the fuzzy random variable is a multi-valued mapping (its values are crisp subsets of \mathbb{R}), the set-valued image $\tilde{X}(\omega)$ restricts mutually exclusive deterministic quantities only (we call these sets *disjunctive* for this reason). The acceptability function then assigns the value 1 to random variables in a certain set of possible ones (the selections of the multi-valued mapping), and the value 0 to the remaining impossible ones. Of course, in the particular case when the fuzzy random variable is a classical random variable (our knowledge its values is represented by sets with only one element) the acceptability function would assign the value 1 to only one random variable, which is the true random variable that models the experiment (and 0 to others). In this case, its observation is completely precise. In [27], Kruse defines the variance of a multi-valued mapping, $\tilde{X} : \Omega \rightarrow \wp(\mathbb{R})$, as the set:

$$\text{Var}_{\text{Kr}}(\tilde{X}) = \{\text{Var}(X) | X \in S(\tilde{X})\},$$

where $S(\tilde{X})$ represents the set of all measurable selections of the multi-valued mapping, i.e., the class of random variables whose values are contained in the (crisp) values of \tilde{X} .

The following example illustrates the merits of this set-valued variance that quantifies the information available about the variance of an underlying random variable, when adopting a “possibilistic” view of the random set, as opposed to the scalar variance of the previous setting. We refer the reader to [44, 45] for additional real life examples.

Example 2.

- (a) The set $\Omega = \{\omega_1, \dots, \omega_4\}$ comprises four objects, whose actual weights are $X_0(\omega_1) = 10.2$, $X_0(\omega_2) = 10.0$, $X_0(\omega_3) = 10.4$, $X_0(\omega_4) = 9.7$.

We sense the weights with a digital device that rounds the measure to the nearest integer, and displays the value ‘10’ in all of these cases. Therefore, we get the constant random set $\Gamma(\omega_i) = [9.5, 10.5], \forall i = 1, \dots, 4$. The true variance of the four measurements is 0.067. Since we only know the information provided by Γ , all we can say about the variance is that it is bounded by the values 0 and 0.25. This is the information that Kruse variance gives us. However, the scalar variance of \tilde{X} returns the misleading value 0.

- (b) Case (a) is an example where the scalar variance of the random set Γ is not an upper bound of the actual value of the variance of X_0 . Neither is it, in general, a lower bound. Indeed, suppose that four objects $\omega_1, \dots, \omega_4$ weigh the same: $X_0(\omega_1) = X_0(\omega_2) = X_0(\omega_3) = X_0(\omega_4) = 9.8g$. Let us also suppose that, for some reason, the weight of the fourth object was imprecisely measured, and we only know that it is between the values 9.5 and 10.5. Our knowledge about the variable X_0 is given by the random set $\Gamma : \Omega \rightarrow \wp(\mathbb{R})$ defined as $\Gamma(\omega_1) = \Gamma(\omega_2) = \Gamma(\omega_3) = \{9.8\}$ and $\Gamma(\omega_4) = [9.5, 10.5]$. The true variance of X_0 is 0, and Kruse variance produces the interval $[0, 0.092]$. But the scalar variance assigns a strictly positive value to it, that depends on the choice of the distance.

The last case suggests that the observed scalar variance of a fuzzy random variable can be misleading in some contexts. It may reflect the variance of the imprecision of the output when all set-realizations are nested (e.g. the knowledge of object ω_4 is more imprecise than the knowledge of the other objects), rather than the actual variability of the underlying phenomenon. In any case, in the above examples, we are not interested in the variability of values of the set-valued-mapping as measured by the scalar variance, but by the possible values of the variance if information had been complete. $\text{Var}_{\text{Kr}}(\tilde{X})$ is a crisp set of potentially attainable variances that reflects the imprecision pervading the observation of the outcome of a random experiment.

This set-valued definition makes full sense when computing the variance of a finite set of incomplete data represented by intervals [26]. Suppose a collection of n measurements of a variable X_0 has been obtained, where each measurement is an interval $[x_i, \bar{x}_i], \forall i = 1, \dots, n$. This interval pertains to a singular measurement step, hence to a precise value x_i of X_0 , that should have been known, had the procedure been perfect. All that is known is that $x_i \in [x_i, \bar{x}_i]$. Then Kruse’s variance (computed for this particular sample) is the set of variances that could have been obtained by substituting each interval $[x_i, \bar{x}_i]$ by the true value of X_0 in realization i . Computing the bounds of the interval $\text{Var}_{\text{Kr}}(\tilde{X})$ is an NP-hard problem, in general [26].

3.2 A fuzzy set-valued variance

The preceding definition can be extended to the case of fuzzy random variables in a natural way. But we must first clarify some ideas about the α -cuts of the fuzzy random variable. Let us consider, for each $\alpha \in [0, 1]$, the multi-valued mapping $\tilde{X}_\alpha : \Omega \rightarrow \wp(\mathbb{R})$ defined as follows:

$$\tilde{X}_\alpha(\omega) = [\tilde{X}(\omega)]_\alpha = \{x \in \mathbb{R} : \tilde{X}(\omega)(x) \geq \alpha\}, \quad \forall \omega \in \Omega. \quad (2)$$

and let $S(\tilde{X}_\alpha)$ be the class of measurable selections of \tilde{X}_α , i.e.,

$$S(\tilde{X}_\alpha) = \{X : \Omega \rightarrow \mathbb{R} \text{ measurable} : X(\omega) \in \tilde{X}_\alpha(\omega), \quad \forall \omega \in \Omega\}.$$

Let us now remember that the acceptability function $\text{acc}_{\tilde{X}}$ represents the possibility distribution associated to a possibility measure, \mathbb{II} , over the class of all the random variables (the measurable mappings from Ω to \mathbb{R}). According to [8, 9], a probability measure \mathbb{P} over the class of all random variables from Ω to \mathbb{R} is dominated by the 2d-order possibility measure \mathbb{II} if and only if it satisfies the inequalities :

$$\mathbb{P}(S(\tilde{X}_\alpha)) \geq 1 - \alpha, \quad \forall \alpha \in [0, 1].$$

Such probability measures form the *credal set* of \mathbb{II} . In words, the possibilistic information provided by the acceptability function is equivalent to the following confidence-level information:

For each $\alpha \in [0, 1]$, the probability that the underlying random variable X_0 belongs to $S(\tilde{X}_\alpha)$ is greater than or equal to $1 - \alpha$.

Having these ideas in mind, the definition of set-valued variance given by Kruse can be naturally extended to fuzzy random variables as follows:

Definition 2. *The fuzzy variance $FVar(\tilde{X})$ of the fuzzy random variable $\tilde{X} : \Omega \rightarrow \mathcal{F}(\mathbb{R})$, is the unique fuzzy set determined by the nested family of sets:*

$$F(\alpha) := FVar(\tilde{X}_\alpha), \quad \forall \alpha,$$

where \tilde{X}_α is the multi-valued mapping α -cut of \tilde{X} .

We refer to the fuzzy set whose membership function is given by the expression

$$\pi_{FVar(\tilde{X})}(s) = \sup\{\alpha \in (0, 1] : s \in \text{Var}_{\text{Kr}}(\tilde{X}_\alpha)\}, \quad \forall x \in \mathbb{R}.$$

Since $\{s : \pi_{FVar(\tilde{X})}(s) > \alpha\} \subseteq F(\alpha) \subseteq \{s : \pi_{FVar(\tilde{X})}(s) \geq \alpha\}, \forall \alpha \in (0, 1)$, it is easy to see that the following (Kwakernaak style) equality holds:

$$\pi_{FVar(\tilde{X})}(s) = \sup\{\text{acc}_{\tilde{X}}(X) : \text{Var}(X) = s\}, \quad \forall s \in \mathbb{R}, \quad (3)$$

where $\text{acc}_{\tilde{X}}(X)$ denotes the acceptability degree of X . The membership degree $\pi_{FVar(\tilde{X})}(s)$ of a value s to the fuzzy set $FVar(\tilde{X})$ represents the maximal possibility degree of random variables whose variance is equal to s .

When the outputs of a random experiment are imprecisely observed, our knowledge about their dispersion is also imprecise. So, the fuzzy variance can be called *potential variance*, since $\pi_{FVar(\tilde{X})}(s)$ is the degree of possibility that s is the variance (in case it exists) of the actual underlying random variable. $FVar(\tilde{X})$ is a fuzzy set of potentially observable variances.

Example 3. *Let us reconsider Example 1, now taking into account the imprecision of the linguistic labels, each label being now viewed as a possibility distribution on the real weight scale. If an apple ω is assigned the linguistic label L_i , it is interpreted as a fuzzy interval $\tilde{X}(\omega)$ restricting the real value $X_0(\omega)$ (the actual weight of the apple), this fuzzy set expressing only the subjective judgment of the grocer. Alternatively we may think of the grocer providing nested subjective prediction intervals $[a_\alpha, b_\alpha]$ on the actual weight, with lower probabilities $1-\alpha$. The second-order model, in this subjectivist perspective addresses the following problem. Given a probability space (Ω, \mathcal{A}, P) and some (non-random) imprecise observation of a quantity (for instance expressed in a linguistic way), what is the family of probabilities P_{X_0} compatible with the imprecision expressed by \tilde{X} ? The natural way of computing the variance here is to use the fuzzy interval $FVar(\tilde{X})$. Assuming apples ω are selected at random (according to the Laplace distribution P) from the container, we can obtain a set of fuzzy evaluations of the apples in the basket: $\{\tilde{X}(\omega) : \omega \in \Omega\}$. Then we can compute the fuzzy variance of the apple weights (reflecting the variability of apple weights only) applying Definition 2.*

The example seems to suggest that the Kruse view of fuzzy random variable should be confined to the case when X_0 is subjectively evaluated. However, the following example suggests it may apply to a wider range of situations where a measurement device is involved.

Example 4. *Let us consider again the situation described in Example 1. But let us now introduce a modification: the grocer now uses scales, but he does not fully trust the obtained measurement. He considers the scales are*

“under control” 90% of the time, and in such situation the measurements are within a 10g error margin. In the remaining 10% of the time, the scales are “out of control” and we can only guarantee an error lower than 50g. Apples are picked at random in a container, so that there is a uniform (Laplace) probability on Ω . Let us denote by $X_0(\omega)$ the ill-known quantity describing the (true) weight of an arbitrary apple $\omega \in \Omega$. Suppose the weight of some apple ω has been measured, and the displayed quantity is $X_e(\omega) = x^*$.

The information of the grocer concerning the imprecision of his scales can be described by a fuzzy interval \tilde{e} , restricting the error e made when measuring apple ω . It corresponds to the possibility distribution

$$\pi_{\tilde{e}}(x) = \begin{cases} 0 & \text{if } x \notin [-50, 50] \\ 0.1 & \text{if } x \in [-50, -10) \cup (10, 50] \\ 1 & \text{if } x \in [-10, 10]. \end{cases} \quad (4)$$

In this example we are interested in representing our knowledge about the true weight of the apple ω , $X_0(\omega)$, based on the displayed quantity $X_e(\omega)$ and the knowledge the grocer has about his scales.

Since $X_0(\omega) + e = X_e(\omega)$, and our imprecise knowledge about e is expressed by the membership function π of \tilde{e} , our imprecise knowledge about the fixed quantity $X_0(\omega)$ can be described by $x^* - \tilde{e}$ (in the sense of fuzzy arithmetics) corresponding to the possibility distribution defined as:

$$\tilde{X}(\omega)(x) = \begin{cases} 0 & \text{if } x \notin [X_e(\omega) - 50, X_e(\omega) + 50] \\ 0.1 & \text{if } x \in [X_e(\omega) - 50, X_e(\omega) - 10) \cup (X_e(\omega) + 10, X_e(\omega) + 50] \\ 1 & \text{if } x \in [X_e(\omega) - 10, X_e(\omega) + 10]. \end{cases} \quad (5)$$

yielding a fuzzy random variable \tilde{X} in the sense of Kruse and Meyer. This possibility distribution reflects the limited confidence of the grocer about his scales. For each pair (ω, x) , $\tilde{X}(\omega)(x)$ represents the grade of possibility that the true weight of the object ω , $X_0(\omega)$, is x . The fuzzy-valued variance will now reflect the grocer’s limited trust on the accuracy of his scales.

In the case of collecting data representing imprecise measurements of variable X_0 , one may have both a most plausible estimate of the real value \bar{x}_i of X_0 for each experiment i and an interval $[a_i, b_i]$ surely containing it. It can be modeled by a triangular fuzzy interval \tilde{X}_i with mode \bar{x}_i and support $[a_i, b_i]$ [2]. $\text{FVar}(\tilde{X})$ is the right concept for computing the empirical fuzzy variance of such a collection of fuzzy realizations. Methods to that effect are devised in [14] for the fuzzy case.

Remark 1. Variance-covariance matrix *Kruse and Meyer's procedure can be applied without too many changes to the variance-covariance matrix. Using similar reasoning methods as above, a possibility distribution over the class of such matrices can be obtained. It models the imprecise knowledge available about the variance-covariance matrix of the "original" underlying random vector. In [34], Meyer proposes a definition of covariance following a path similar to Kruse's. As often when handling interval-valued information, the combination of the information provided separately about the variance of every component and about the covariance between them is more imprecise than the straight information about the variance-covariance matrix.*

3.3 Observable vs. potential variance

When the fuzzy random variable (or the random set) represents the imprecise observation of a standard random variable, the description of the changes of the observed sets or fuzzy sets via a "classical" variance is not enough to inform about the variability of the underlying phenomenon, as we have shown in Examples 2, 3, 4. In fact, in the context of an imprecisely observed random variable, the scalar variance of Section 2, when non-zero, may only account for an "observable variance", namely the part of the variance that can be measured, despite the imprecision of the observation. It may sometimes partially account for the variability of the underlying phenomenon. For instance, if the fuzzy random variable represents an imprecisely observed random variable with disjoint imprecise realizations, then it has a positive scalar variance that reveals the non-deterministic nature of the underlying process (even if only partially).

On the other hand, as the above examples show, a zero scalar variance is not enough to conclude whether the observed phenomenon is random or not. Example 2(a) leads to a zero observable variance, because the variability of the weight is drowned in the imprecision of the observation. Likewise, the Lubiano et. al. scalar variance applied to Example 4 would provide a debatable estimation of the real variance of X_0 because, in this particular case involving symmetric fuzzy intervals that are identical up to a translation, it would come down to assuming that its realizations lie at the midpoint $X_e(\omega)$ of the core of the fuzzy observations. Moreover, a positive scalar variance cannot reveal the actual randomness of the phenomenon if the realizations are nested fuzzy sets. It only points out the variability of the imprecision of the observed outcomes, as does the Körner scalar variance in Example 2(b).

In fact, one way of computing the observable variance as a scalar is to choose an appropriate distance d_{\min} between fuzzy sets in the scalar variance (1), namely one that vanishes when the two fuzzy intervals overlap: consider

two fuzzy intervals F and G , and let

$$d_{\min}(F_\alpha, G_\alpha) = \inf\{|x - y|, x \in F_\alpha, y \in G_\alpha\},$$

and (for instance) $d_{\min}(F, G) = \inf_{\alpha > 0} d_{\min}(F_\alpha, G_\alpha)$. We can check that this new scalar variance is less than the lower bound of Kruse variance. In Example 2(b), the above scalar variance is now 0.

4 Interval-valued variance associated to a first-order imprecise model

The interval-valued variance now introduced in this section is based on the first order imprecise model presented in [1, 5, 10]. Suppose, as in the previous example, that there is an imperfect measurement process that makes the observation of each value $X_0(\omega)$ noisy. So there is a sequence of two random experiments whose sample spaces are Ω and \mathbb{R} , respectively. Let us suppose, on the one hand, that the probability distribution on $(\Omega, \mathcal{A}, P) : \mathcal{A} \rightarrow [0, 1]$, is completely determined (in the preceding expression, \mathcal{A} denotes a σ -algebra of events over Ω .) On the other hand, the other experiment is only known via a fuzzy relation which assigns, to each outcome ω of the first sub-experiment in the sample set Ω , the fuzzy set $\tilde{X}(\omega)$ of possible outcomes of the second experiment. Its membership function is a possibility distribution $\tilde{X}(\omega)(\cdot) = \pi(\cdot|\omega)$ that models knowledge about the relationship between the outcome ω of the first sub-experiment and the possible outcomes of the second one. $\pi(\cdot|\omega)$ is called a conditional possibility distribution : if the result of the first experiment is ω , then the possibility degree of $x \in B$ occurring in the second one is $\Pi(B|\omega) = \sup_{x \in B} \tilde{X}(\omega)(x)$. The family of conditional probability measures $Q(\cdot|\omega)$ (indexed by ω) encoded by $\Pi(B|\omega)$ is denoted by $\mathcal{C} = \{Q(\cdot|\omega) : Q(B|\omega) \leq \Pi(B|\omega) \forall B \in \beta_{\mathbb{R}}, \omega \in \Omega\}$.

As a consequence, we have partial information about the probability distribution P_2 on $(\mathbb{R}, \beta_{\mathbb{R}})$, resulting from the sequence of two random experiments. The combination of both sources of information using natural extension techniques [47] allows to describe the available information about this probability distribution by means of an upper probability. In fact, this probability measure is given by the formula:

$$P_2(B) = \int_{\Omega} Q(B|\omega) dP(\omega), \text{ where}$$

$$Q(B|\omega) \leq \Pi(B|\omega), \forall \omega \in \Omega, \forall B \in \beta_{\mathbb{R}}.$$

The interval-valued variance is then naturally defined as follows:

Definition 3. Consider a probability space (Ω, \mathcal{A}, P) , and a fuzzy random variable defined over it, $\tilde{X} : \Omega \rightarrow \mathcal{F}(\mathbb{R})$. For each $\omega \in \Omega$, let $\Pi(\cdot|\omega)$ denote the possibility measure associated to the possibility distribution $\tilde{X}(\omega)$. We define the first-order imprecise variance of \tilde{X} as the (crisp) set $\text{IVar}(\tilde{X}) =$

$$\left\{ \text{Var}(Q_2) : Q_2(A) = \int_{\Omega} Q(A|\omega) dP(\omega), \forall A \in \beta_{\mathbb{R}} \text{ s.t. } Q(B|\omega) \leq \Pi(B|\omega) \forall B \in \beta_{\mathbb{R}} \right\}.$$

$\text{IVar}(\tilde{X})$ reflects the set of possible values of the variance of the second sub-experiment, according to the available information.

Despite the fact that the present model is, like the one presented in the previous section, associated to a possibilistic interpretation of fuzzy sets, the meaning of the two definitions of variance derived from them are quite different. In the previous section, the fuzzy random variable, \tilde{X} , represented an imprecise observation of a particular (classical) random variable, X_0 . For each possible result of the random experiment, $\omega \in \Omega$, the value $X_0(\omega)$ is fixed but we have imprecise knowledge about it. However, in the present model, the fuzzy random variable \tilde{X} represents our (imprecise) knowledge about the link between two steps of a random experiment. Thus, the same outcome ω in the first step can be associated to different outcomes of the second step. Under the above assumptions, we must combine the probability measure associated to the first step with the probability measure that relates the first step with the second one. As our knowledge about the latter conditional probability measure is given by a credal set of probabilities, so is our knowledge about the probability measure that governs the whole process. Thus, our knowledge about the variability of the results of the second sub-experiment will be characterized by a pair of lower and upper bounds.

It is obvious that this is a generalization of the concept of variance of a classical random variable: in fact, let us suppose that the values of the fuzzy random variable \tilde{X} are real values. In other words, suppose that for all $\omega \in \Omega$, $\Pi(\cdot|\omega)$ is, in particular, the degenerated probability measure in a point $X(\omega)$. It is easy to prove that the class of probability functions $\mathcal{D} = \{\int_{\Omega} Q(\cdot|\omega) dP(\omega), \forall Q(\cdot|\omega) \leq \Pi(\cdot|\omega)\}$, is reduced to the singleton $\{P_X\}$ (it is the only probability measure compatible with P and $\Pi(\cdot|\cdot)$).

Example 5. Let us turn back to the situation described in Example 4. We assume that the observed value on the scale, say $X_E(\omega)$, is noisy for a fixed apple ω . If we choose the same apple 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 random quantity $X_E(\omega)$, regardless of having performed a measurement or not. This is different from the previous example where we observed the measured value (the

estimated weight of an apple) and described our knowledge about the true weight $X_0(\omega)$. Here we describe our knowledge about the measured value, based on randomly picking an apple, and measuring its weight with noisy scales. The variance of X_E is then a function of the variance of X_0 and the variance of the random error of the scales.

Namely, suppose

$$X_E(\omega) = X_0(\omega) + E,$$

where E is a random error due to the variability of the scales. If the probability measure of the error is P_E , then the conditional probability of the random displayed quantity $X_E(\omega)$ belonging to some $A \subseteq \mathbb{R}$, for an apple ω , is

$$P(X_E \in A|\omega) = P_E(\{x - X_0(\omega) : x \in A\}).$$

The probability distribution P_{X_E} is of the form

$$P_{X_E}(A) = \sum_{\omega \in \Omega} P(X_E \in A|\omega) \cdot P(\{\omega\}).$$

Unfortunately, the knowledge about P_E is imprecise. All we know about it is that

$$P_E([-10, 10]) \geq 0.9 \text{ and } P_E([-50, 50]) = 1.$$

According to Dubois and Prade [16] and Couso et al. [7], those constraints are equivalently expressed by means of a fuzzy interval \tilde{e} with possibility distribution $\pi_{\tilde{e}}(e) = 1$ if $e \in [-10, 10]$, and 0.1 if $e \in [-50, -100] \cup [10, 50]$, and 0 otherwise (this is the same possibility distribution $\pi_{\tilde{e}}$ defined in Example 4, now representing imprecise knowledge about a random phenomenon). It also corresponds to the basic mass assignment m :

$$m([-10, 10]) = 0.9 \text{ and } m([-50, 50]) = 0.1,$$

where 0.9 and 0.1 respectively represent the probabilities of being under/out of control, and the focal sets $[-10, 10]$ and $[-50, 50]$ represent sets of values among which the probabilities 0.9 and 0.1 can be shared. Our knowledge about the conditional probability $P(X_E \in A|\omega)$ is then of the form

$$P(X_E \in A|\omega) \leq \Pi(A|\omega) = \sup_{x \in A} \tilde{X}_E(\omega)(x).$$

The probability function P_{X_E} is thus expressed by a mass function m_{X_E} . This mass function has focal sets of the form $[X_0(\omega) - 10, X_0(\omega) + 10]$ with

probability $0.9P(\{\omega\})$ and $[X_0(\omega)-50, X_0(\omega)+50]$ with probability $0.1P(\{\omega\})$ ^{4,5}. The set of probabilities it encompasses reflects not only the random picking process in Ω , but also the random error tainting the measurement process. Hence the variance of the fuzzy random set corresponding to this problem is the range of the variances of all probability distributions in the probability family characterized by the mass assignment m_{X_E} .

The difference with Example 4 is patent: with the latter we try to represent our knowledge about $X_0(\omega)$ after observing $X_E(\omega) = x$ thus not considering the added variability due to the error, just the grocer's lack of knowledge about it; here we represent our knowledge about the possible observed measured apple weight that cumulate variability due to the measurement process and the variability due to the choice of the apple.

5 Fuzzy versus interval-valued variance

In the last two sections, we have presented situations where each non-scalar variance should be used. We have shown that the fuzzy variance in Kruse-Meyer's framework is natural when the goal is to describe the imprecise information about the dispersion of an underlying classical random variable. On the other hand, the interval-valued variance is appropriate when the goal is to represent the range of the variance in a two-stepped random experiment, if the fuzzy random variable is interpreted as a set of conditional probabilities encoded by a conditional possibility measure. But we may try to investigate the relationships between both definitions. On the one hand, we observe that, when the fuzzy random variable reduces to a random set, the fuzzy variance produces a crisp set of numbers. So we may ask ourselves whether the latter is somehow related to the interval-valued variance of the random set. On the other hand, intervals are easier to manage than fuzzy intervals, when they are used in a decision process. Thus, it would be interesting to investigate whether the interval-valued variance can be interpreted in Kruse & Meyer's context. In the following subsections, we deal with these two issues.

⁴Because the probability on Ω is uniform, this notation is a bit unusual in the sense that we may have the same interval-valued focal set for different ω 's. In this case the convention is to add the weights pertaining to the same focal set.

⁵Note that these focal sets are not the realizations of a random process, since when measuring the weight of an apple only a precise value $x^* = X_e(\omega)$ for a prescribed apple ω and a prescribed error $E = e$, will be read on the scales. These focal sets only express the available knowledge about the random "picking + measurement" process.

5.1 The particular case of random sets

When the fuzzy random variable is a random set $\tilde{X} : \Omega \rightarrow \wp(\mathbb{R})$, the fuzzy variance coincides with Kruse variance, $\text{FVar}(\tilde{X}) = \{\text{Var}(X) : X \in S(\tilde{X})\}$. Let us note that, for each measurable selection $X \in S(\tilde{X})$, the probability measure P_X satisfies the following restrictions: $\forall A \in \beta_{\mathbb{R}}$,

$$\underline{P}_{\tilde{X}}(A) = P(\{\omega : \tilde{X}(\omega) \subseteq A\}) \leq P_X(A) \leq \overline{P}_{\tilde{X}}(A) = P(\{\omega : \tilde{X}(\omega) \cap A \neq \emptyset\}),$$

where $\underline{P}_{\tilde{X}}$ and $\overline{P}_{\tilde{X}}$ respectively denote the Dempster lower and upper probabilities associated to the random set \tilde{X} (assuming that \tilde{X} is strongly measurable and $\tilde{X}(\omega)$ is never empty). Let $\Pi(\cdot|\omega)$ be the Boolean conditional possibility measure encoding $\tilde{X}(\omega)$. It is checked in [41] that the class of probabilities $\mathcal{D} = \{\int_{\Omega} Q(\cdot|\omega) dP(\omega), \forall Q(\cdot|\omega) \leq \Pi(\cdot|\omega)\}$ induced by an ill-known conditional probability dominated by a Boolean possibility measure coincides with the class of probability measures that are dominated by the Dempster upper probability. In other words, the following inclusion holds:

$$\mathcal{P}(\tilde{X}) = \{P_X : X \in S(\tilde{X})\} \subseteq \mathcal{D} = \{Q \leq \overline{P}_{\tilde{X}}\}.$$

The converse inclusion does not hold in general, as it is checked in [11, 35, 36, 39, 40]. Since Kruse variance is the class of classical variances induced by $\mathcal{P}(\tilde{X})$ and the interval-valued variance is induced by \mathcal{D} , we conclude that the Kruse variance of a random set \tilde{X} is a set contained in the interval-valued variance: $\text{FVar}(\tilde{X}) \subseteq \text{IVar}(\tilde{X})$. Let us give an example to clarify this idea:

Example 6. Let Ω be a singleton $\{\omega\}$ and let $\tilde{X} : \Omega \rightarrow \wp(\mathbb{R})$ be the random set defined as $\tilde{X}(\omega) = [a, b]$, where $[a, b]$ denotes an arbitrary interval in the real line (We assume that $a < b$). On the one hand, we can check that

$$\mathcal{P}(\tilde{X}) = \{P_X : X \in S(\tilde{X})\} = \{D_c : c \in [a, b]\},$$

where D_c denotes the Dirac probability measure degenerated on c , i.e.:

$$D_c(A) = \begin{cases} 1 & \text{if } A \ni c \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, we can check that $\mathcal{D} = \{Q : Q([a, b]) = 1\}$. Thus, the class of probability measures associated to the random selections of \tilde{X} is a non-convex set strictly included in \mathcal{D} . Such absence of convexity has an important impact in the calculation of the variance. The fuzzy variance of \tilde{X} (Kruse's variance) is

$$\text{FVar}(\tilde{X}) = \{\text{Var}(X) : X \in S(\tilde{X})\} = \{0\}.$$

On the other hand, the interval-valued variance is

$$\text{IVar}(\tilde{X}) = \{\text{Var}(Q) : Q([a, b]) = 1\} = \left[0, \frac{(b-a)^2}{4}\right].$$

For each $p \in (0, 1)$, the probability measures Q_p satisfying $Q_p(\{a\}) = p$, $Q_p(\{b\}) = 1 - p$ belong to \mathcal{D} but not to $\mathcal{P}(\tilde{X})$. The upper bound of $\text{IVar}(\tilde{X})$ is reached when $p = \frac{1}{2}$. The variance of a linear convex combination of two probability measures is strictly greater than the convex combination of the variances, except when they have the same expectation. Furthermore, according to [10], the convex hull of the class $\mathcal{P}(\tilde{X})$ is always contained in the class \mathcal{D} , used to compute the interval variance. This is the reason why the upper bound of the interval-valued variance may be strictly greater than the upper bound of Kruse's variance, when \tilde{X} is a random set.

The last example shows that the fuzzy variance generally does not coincide with the interval-valued variance, when \tilde{X} is a random set. When calculating the fuzzy variance, we are assuming that the final outcome in \mathbb{R} is univocally determined by the outcome in Ω , $Q(\cdot|\omega)$ being degenerated. Such a restriction on the conditional probability measure $Q(\cdot|\omega)$ is not assumed in the calculation of the interval-valued variance.

5.2 Reducing the fuzzy variance to the interval-valued one

The definition of the fuzzy variance FVar of a general fuzzy random variable is based on the concept of acceptability function introduced at the beginning of Section 3, which is a possibility distribution over the class of all measurable mappings from Ω to \mathbb{R} . Based on this concept, we have defined the fuzzy variance of \tilde{X} as the fuzzy set associated to the possibility distribution $\pi_{\text{FVar}} : \mathbb{R}^+ \rightarrow [0, 1]$ defined by Equation (3). The membership value $\pi_{\text{FVar}(\tilde{X})}(s)$ represents the possibility grade that the variance of the underlying random variable is equal to s . In a similar way, we can [8, 9] construct a possibility measure over the class of all probability measures on \mathbb{R} . It represents imprecise information about the probability measure P_{X_0} induced by the underlying random variable. This “second-order” possibility distribution, π , is defined as follows:

$$\pi(Q) = \sup\{\text{acc}_{\tilde{X}}(X) : P_X = Q\}.$$

The quantity $\pi(Q)$ represents the possibility grade that the probability measure induced by X_0 , namely P_{X_0} , coincides with Q .

According to a procedure developed by Walley [48], any second-order possibility measure can be reduced into a pair of (first-order) upper and lower probabilities. Let us briefly describe Walley's reduction in our context. Let \mathbb{I} denote the second-order possibility measure associated to π . Let us first notice that it is an upper probability over the class $\mathbb{P}\mathbb{R}(\Omega)$ of standard probabilities, so it encodes a second-order credal set, i.e. a class of second-order probability measures, $\mathcal{P} = \{\mathbb{P} : \mathbb{P}(\mathcal{Q}) \leq \mathbb{I}(\mathcal{Q}), \forall \mathcal{Q} \subseteq \mathbb{P}\mathbb{R}(\Omega)\}$, where \mathbb{I} is the possibility measure induced by π on the set $\mathbb{P}\mathbb{R}(\Omega)$ of probability measures over Ω , and \mathcal{Q} is a measurable set of probability measures. In other words,

$$\mathbb{I}(\mathcal{Q}) = \sup\{\mathbb{P}(\mathcal{Q}) : \mathbb{P} \in \mathcal{P}\}, \forall \mathcal{Q} \subseteq \mathbb{P}\mathbb{R}(\Omega).$$

In this setting, Walley derives a standard probability measure $P_{\mathbb{P}}$ from each of the above second-order probability measures \mathbb{P} . To give a flavor of Walley's procedure, let us assume, for the sake of simplicity, that π assigns non-zero possibilities only to a finite class of first-order probability measures $\{Q_1, \dots, Q_n\}$ and let A denote an arbitrary measurable event. In Walley's context, fixing a second-order probability \mathbb{P} , the probability $Q_j(A)$ is understood as the following conditional probability:

If the probability measure that governs the experiment on \mathbb{R} is Q_j , then the probability of occurrence of the event A is $Q_j(A)$.

Then, according to the Total Probability Theorem, Walley calculates the probability of occurrence of the event A as the following average:

$$P_{\mathbb{P}}(A) = \sum_{j=1}^n Q_j(A) \cdot \mathbb{P}(Q_j), \forall A \in \beta_{\mathbb{R}},$$

where $\mathbb{P}(Q_j)$ is the probability that the correct probability measure is Q_j . Walley derives, in a natural way, a pair of upper and lower probabilities from \mathbb{I} as follows:

$$\underline{P}_W(A) = \inf_{\mathbb{P} \in \mathcal{P}} P_{\mathbb{P}}(A) \text{ and } \overline{P}_W(A) = \sup_{\mathbb{P} \in \mathcal{P}} P_{\mathbb{P}}(A), \forall A \in \beta_{\mathbb{R}}.$$

The above lower and upper probabilities can be understood as lower and upper bounds for the probability of occurrence of A . But mind that two different types of probabilities are combined within the same formula. In fact, the first order probabilities Q_j are objective and express the variability of occurrence of events, while second-order probabilities \mathbb{P} are subjective and qualify our confidence in first-order probability statements. Nevertheless, following Kyburg [30], both of them can be combined into a joint probability

space, even if they are conceptually different kinds of probabilities, whenever a concrete application requires a “single-case” probability.

As a consequence, the second-order possibility measure associated to the fuzzy random variable \tilde{X} , when Kruse & Meyer’s approach is followed, can be reduced into a pair of upper and lower probabilities. Furthermore, it is checked in [10] that the upper and lower probabilities associated to such a reduction coincide with the upper and lower bounds induced by the class of probability measures $\mathcal{D} = \{\int_{\Omega} Q(\cdot|\omega) dP(\omega), \forall Q(\cdot|\omega) \leq \Pi(\cdot|\omega)\}$. Specifically, let \underline{P} and \overline{P} denote:

$$\underline{P}(A) = \inf_{P_2 \in \mathcal{D}} P_2(A) \text{ and } \overline{P}(A) = \sup_{P_2 \in \mathcal{D}} P_2(A).$$

They represent the (first-order) upper-lower model considered in the definition of the interval-valued variance. Then, according to Theorem 7.3 in [10], the following equalities hold:

$$\underline{P}_W(A) = \underline{P}(A), \text{ and } \overline{P}_W(A) = \overline{P}(A), \forall A \in \beta_{\mathbb{R}},$$

when $\tilde{X}_{\alpha}(\omega)$ is an open or a closed set, for each $\alpha \in [0, 1]$ and each $\omega \in \Omega$. These equalities allow us to reinterpret the interval-valued variance within Kruse and Meyer’s context, since it comes down to generating the convex hull of the probability set at work in the second approach.

Other possible reduction techniques exist in the literature:

- Walley [48] considers the cuts of the second-order possibility distribution π induced from the acceptability function as a random set of probability measures, namely a nested set of probability families $\{\mathcal{Q}_{\alpha}, \alpha \in (0, 1]\}$ where $\mathcal{Q}_{\alpha} \subseteq \mathbb{PR}(\Omega)$ is the α -cut of π . Let $P_{\alpha}^*(A) = \sup\{P(A), P \in \mathcal{Q}_{\alpha}\}$, for $A \in \beta_{\mathbb{R}}$. The following upper probability function P^* is derived as

$$P^*(A) = \int_0^1 P_{\alpha}^*(A) d\alpha.$$

Walley proved that P^* is the same upper probability as \overline{P}_W derived above.

- The straightforward reduction of a random fuzzy set proposed in [1, 3] looks different from Walley’s procedures but turn out to yield the same results. Namely, consider a fuzzy random variable \tilde{X} from (Ω, \mathcal{A}, P) to \mathbb{R} as a probability distribution Q over a set of fuzzy intervals \tilde{x}_i inducing possibility measures $\{\Pi_1, \dots, \Pi_k\}$. This is done by letting $Q(\{\tilde{x}_i\}) =$

$P(\{\omega : \tilde{X}(\omega) = \tilde{x}_i\})$ (assuming a finite setting for simplicity). It can be reduced to a probability family generated by the random set associated to the Shafer plausibility function

$$Pl = \sum_{j=1}^k \Pi_j \cdot Q(\{\tilde{x}_j\}).$$

Couso and Sánchez [10] recently proved (in a more general setting) that this function actually coincides with P^* (hence with \bar{P}_W), which is thus a Shafer plausibility function. This reduction from a fuzzy random variable to a belief function and the ensuing interval-valued variance can thus be considered as canonical.

Summarizing,

- On the one hand, the fuzzy variance represents a possibility distribution over the class of possible values for the variance of the underlying random variable, when we follow Kruse & Meyer approach. It is the kind of variance that we should use when “objective probabilities about the occurrence of events” are distinguished from “belief degrees about the values of the probability of events”, and both types of uncertainty stay at two different levels.
- On the other hand, the interval-valued variance admits of two different interpretations:
 - (a) It can be viewed as the range of the values for the dispersion of the outcomes due to an ill-known second random sub-experiment, modelled as a family of conditional probabilities encoded by the fuzzy random variable.
 - (b) It can be also understood as the range of the values of the variance of a probability family obtained by defuzzifying the higher-order possibility distribution over probabilities present in Kruse-Meyer view, and convexifying the resulting probability set.

The above discussion can be illustrated on Examples 4 and 5.

- The fuzzy variance of \tilde{X} in Example 4 determines a possibility distribution on the real line. The membership value of some variance s represents the grade of possibility that the variance of the (true) weights of the apples in the container is equal to s . It does not account for the measurement process variability.

- The interval-valued variance of \tilde{X}_E Example 5 represents, in a natural way, the range of possible values for the variance of the numbers displayed by the scale, if we allow repeated measurements on the same apple. It combines two sources of variability due to the picking procedure in the apple container, and the measurement process, respectively. We know with certainty that such a variance belongs to such an interval.

But now, according to the ideas described in this section, we can also give a meaningful interpretation to an interval-valued variance of \tilde{X} in Example 4, derived as explained in this section. In fact, this interval-valued variance combines two sources of dispersion: objective dispersion due to randomness about the choice of the apple and potential dispersion tied to imprecision in knowledge. Note that the latter dispersion is potential: it is accounted for “as if” the lack of specificity regarding the knowledge of X_0 were due to an underlying ill-known additional random process, a feature that may or not agree with the particulars of the considered application, and that makes the obtained interval-valued variance wider than in the Kruse-Mayer original view.

6 Concluding remarks

This paper has laid bare three different views of fuzzy random variables that underlie different intuitions and lead to different extensions of the concept of variance. None of them looks, in general terms, preferable to the others, but they serve different purposes: they reflect different models of the observed phenomenon, as well as different assumptions on the available knowledge about this phenomenon. According to the problem under concern, it should be decided whether the dispersion needs to be measured as a number, a fuzzy set or a crisp set. The scalar variance measures the variability of the membership function, considered as an observable entity, not the variability of an underlying precise quantity it may possibly describe. This definition is in the spirit of classical statistics and it does not take into account the idea that the values of a fuzzy random variable model incomplete knowledge. Some proposals for scalar variance are equivalent to considering first a representative (numerical) element of every fuzzy realization of the fuzzy random variable (the midpoint of the support, for instance) and then calculate the dispersion of these numerical values. Part of the actual variability can be observed and measured by means of scalar variance if the fuzzy outcomes are precise enough and/or often disjoint. On the other hand, the average

precision of the fuzzy random variable, and the variance of this precision are other useful evaluations.

On the contrary, interval-valued and fuzzy variances measure the potential variability of an ill-known random variable. If the fuzzy random variable represents the ill-known measurement of some characteristics of a classical random process, one of the two non-scalar definitions must be used.

- In the Kruse-Meyer view, a fuzzy random variable is a fuzzy set of possible classical random variables. It represents an ill-observed random variable. The fuzzy values of the random variable represent nested confidence intervals on the actual outcome of each trial. The fuzzy variance represents knowledge, in the form of induced nested confidence intervals, about the true dispersion of the random variable under study, due to the variability inside the sample space.
- In the third view, the fuzzy random variable is viewed as a fuzzy relation between the sample space and the range of the random variable, encoding a family of conditional probabilities, each of which potentially representing a random measurement process. Hence a family of probability measures on the range of the random variable is obtained, leading to an interval representing all potential values of the actual variance. This interval variance represents the set of all possible values for the dispersion of a random variable and its associated ill-known random measurement process, thus combining two sources of variability.

Acknowledgements

This work has been supported by grant TIN2007-67418-C03-03, and by the ANR project CRISCO2.

References

- [1] C. Baudrit, I. Couso, D. Dubois (2007) Joint propagation of probability and possibility in risk analysis: Towards a formal framework, *Int. J. of Approximate Reasoning* **45** 82-105.
- [2] C. Baudrit, D. Dubois (2006) Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis* 51(1): 86-108.

- [3] C. Baudrit, D. Dubois, D. Guyonnet, H. Fargier (2006) Joint treatment of imprecision and randomness in uncertainty propagation. In: *Modern Information Processing: From Theory to Applications*. (eds: B. Bouchon-Meunier, G. Coletti, R.R. Yager), Elsevier, 37-47.
- [4] C. Bertoluzza, A. Salas, N. Corral (1995) On a new class of distances between fuzzy numbers, *Mathware and Soft Computing* **2** 71-84.
- [5] I. Couso, E. Miranda, G. de Cooman (2004) A possibilistic interpretation of the expectation of a fuzzy random variable. In *Soft methodology and random information systems* (eds: M. López-Díaz, M. A. Gil; P. Grzegorzewski, O. Hryniewicz, and J. Lawry), 133-140. Springer, Heidelberg (Germany).
- [6] I. Couso, S. Montes, P. Gil (1998) Función de distribución y mediana de variables aleatorias difusas, *Proceedings of the Conference ESTYLF'98* (Pamplona, Spain).
- [7] I. Couso, S. Montes, P. Gil (2001) The necessity of the strong alpha-cuts of a fuzzy set, *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems* **9** 249-262.
- [8] I. Couso, S. Montes, P. Gil (2002) Second-order possibility measure induced by a fuzzy random variable. In *Statistical modeling, analysis and management of fuzzy data* (eds: C. Bertoluzza, M. A. Gil and D. A. Ralescu), 127-144. Physica-Verlag, Heidelberg.
- [9] I. Couso, L. Sánchez (2008) Higher order models for fuzzy random variables, *Fuzzy Sets and Systems* **159** 237-258.
- [10] I. Couso, L. Sánchez (2008) Upper and lower probabilities induced by a fuzzy random variable, submitted to *Fuzzy Sets and Systems*.
- [11] I. Couso, L. Sánchez, P. Gil (2004) Imprecise distribution function associated to a random set, *Information Sciences* **159** 109-123.
- [12] G. de Cooman, P. Walley (2002) An imprecise hierarchical model for behaviour under uncertainty, *Theory and Decision* **52** 327-374.
- [13] P. Diamond, P. Kloeden (1994) *Metric Spaces of Fuzzy Sets*, World Scientific, Singapur.
- [14] D. Dubois, H. Fargier, J. Fortin (2005) The empirical variance of a set of fuzzy intervals. *Proc. IEEE Int. Conf. on Fuzzy Systems*, Reno, Nevada, IEEE Press, 885-890.

- [15] D. Dubois, H. Prade (1987) The mean value of a fuzzy number, *Fuzzy Sets and Systems* **24** 279-300.
- [16] D. Dubois, H. Prade (1992) When upper probabilities are possibility measures, *Fuzzy Sets and Systems* **49** 1992, 65-74.
- [17] D. Dubois, H. Prade (1997) The three semantics of fuzzy sets, *Fuzzy Sets and Systems* **90** 141-150.
- [18] Y. Feng, L. Hu, H. Shu (2001) The variance and covariance of fuzzy random variables and their applications, *Fuzzy Sets and Systems* **120** 487-497.
- [19] R. Féron (1976) Ensembles aléatoires flous, *C.R. Acad. Sci. Paris Ser. A* **282** 903-906.
- [20] M. Fréchet (1948) Les éléments aleatoires de natures quelconque dans un espace distancié, *Ann. Inst. H. Poincaré* **10** 215-310.
- [21] E.P. Klement, M.L. Puri, D.A. Ralescu (1986) Limit theorems for fuzzy random variables, *Proc. Roy. Soc. London A* **407** 171-182.
- [22] R. Körner (1997) On the variance of fuzzy random variables, *Fuzzy Sets and Systems* **92** 83-93.
- [23] V. Krätschmer (2001) A unified approach to fuzzy random variables, *Fuzzy Sets and Systems* **123** 1-9.
- [24] V. Krätschmer (2002) Some complete metrics on spaces of fuzzy subsets *Fuzzy Sets and Systems*, **130**, 357-365
- [25] V. Krätschmer (2006) Integrals of random fuzzy sets, *Test* **15** 433-469.
- [26] V. Kreinovich, G. Xiang and S. Ferson (2006) Computing mean and variance under Dempster-Shafer uncertainty: Towards faster algorithms *Int.J. of Approximate Reasoning*, 42(3), 212-227.
- [27] R. Kruse (1987) On the variance of random sets, *J. Math. Anal. Appl.* **122** 469-473.
- [28] R. Kruse, K.D. Meyer (1987) *Statistics with vague data*. D. Reidel Publishing Company.
- [29] Kwakernaak (1989) Fuzzy random variables. Definition and theorems, *Inform. Sci.* **15** 1-29.

- [30] H.E. Kyburg (1988) Higher order probabilities and intervals, *Int. J. of Approximate Reasoning* **2** 195-209.
- [31] Y.K. Liu, B. Liu (2003) A class of fuzzy random optimization: expected valued models, *Information Sciences* **155** 89-102.
- [32] M.A. Lubiano, Variation measures for imprecise random elements, Ph.D. Thesis, University of Oviedo (Spain, 1999).
- [33] M.A. Lubiano, M.A. Gil (2002) f -inequality indices for fuzzy random variables, In *Statistical modeling, analysis and management of fuzzy data*. C. Bertoluzza, M. A. Gil and D. A. Ralescu (Eds.), 43-63. Physica-Verlag, Heidelberg.
- [34] K.D. Meyer, R. Kruse (1990) On calculating the covariance in the presence of vague data. In: *Progress in Fuzzy Sets and Systems*. W.H. Janko, M.Roubens and H.J. Zimmermann (Eds.) Kluwer Academic Publishers, Dordrecht.
- [35] E. Miranda, I. Couso, P. Gil (2002) Upper probabilities and selectors of random sets. In: *Soft methods in probability, statistics and data analysis* P. Gzegorzewski, O. Hryniewicz, M.A. Gil (Eds.). Physica-Verlag, Heidelberg (Germany).
- [36] E. Miranda, I. Couso, P. Gil (2003) Study of the probabilistic information of a random set, *Proceedings of the 3rd ISIPTA Conference*. Lugano (Switzerland).
- [37] E. Miranda, I. Couso and P. Gil (2002) Relationships between possibility measures and nested random sets, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10** 1-15.
- [38] E. Miranda, I. Couso and P. Gil (2004) A random set characterisation of possibility measures, *Information Sciences* **168** 51-75.
- [39] E. Miranda, I. Couso, P. Gil (2005) Random sets as imprecise random variables. *Journal of Mathematical Analysis and Applications* **307** 32-47.
- [40] E. Miranda, I. Couso, P. Gil (2005) Random intervals as a model for imprecise information, *Fuzzy Sets and Systems* **154** 386-412.
- [41] E. Miranda, G. de Cooman, I. Couso (2005) Imprecise probabilities induced by multi-valued mappings, *J. Stat. Plann. Inference* **133** 173-197.

- [42] R. Moore, Methods and applications of interval analysis. SIAM Studies in Applied Mathematics, 1979.
- [43] M.L. Puri, D. Ralescu (1986) Fuzzy Random Variables, *J. Math. Anal. Appl.* **114** 409-422.
- [44] L. Sánchez, I. Couso (2007) Advocating the Use of Imprecisely Observed Data in Genetic Fuzzy Systems, *IEEE Transactions on Fuzzy Systems* **15** 551-562.
- [45] L. Sánchez, I. Couso, J. Casillas (2008) Genetic Learning of Fuzzy Rules based on Low Quality Data, Submitted to Fuzzy Sets and Systems.
- [46] G. L.S. Shackle (1961) *Decision, Order and Time in Human Affairs*, (2nd edition), Cambridge University Press, UK.
- [47] P. Walley (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London (UK).
- [48] P. Walley (1997) Statistical inferences based on a second-order possibility distribution, *International Journal of General Systems* **26** 337-384.
- [49] L. A. Zadeh (1975) The concept of a linguistic variable and its application to approximate reasoning, Part 1 *Information Sciences* **8** 199-249.
- [50] L. A. Zadeh (1978) Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* **1** 3-28.