On various definitions of the variance of a fuzzy random variable

Inés Couso
Dep. Statistics and O.R.
Univ. of Oviedo, Spain
couso@uniovi.es

Didier Dubois
IRIT-CNRS
Univ. P. Sabatier, France
Didier.Dubois@irit.fr

Susana Montes
Dep. Statistics and O.R.
Univ. of Oviedo, Spain
montes@uniovi.es

Luciano Sánchez
Dep. Computer Sciences
Univ. of Oviedo, Spain
luciano@uniovi.es

Abstract

According to the current literature, there are two different approaches to the definition of the variance of a fuzzy random variable. In the first one, 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 the second case, the variance of the fuzzy random variable is defined as a crisp number, that makes it easier to take some further decisions. In this work, we introduce a new definition of the variance of a fuzzy random variable, in the context of the theory of imprecise probabilities. The new variance is not defined as a fuzzy or crisp number, but it is a real interval, which is a compromise between both definitions. Our main objectives are twofold: first, we show the interpretation of the new variance and, second, with the help of simple examples, we demonstrate the usefulness of all these definitions when applied to particular situations.

Keywords: Fuzzy random variable, variance, imprecise probabilities.

1 Introduction

The concept of fuzzy random variable, that extends the classical definition of random variable, was introduced by Féron [14] in 1976, and modified by other authors as Kwakernaak [23], Puri and Ralescu [32], Kruse and Meyer [**?**], or Diamond and Kloeden [9], among others. In [18], Krätschmer surveys all of these definitions and proposes an unified approach. In all of these works, 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, many works have generalized different concepts and classical results to the case in which all observations associated to the different results of the experiment are fuzzy sets.

Regarding the generalizations in this context of the definitions of the parameters associated to a probability distribution, we can divide them into two groups. On the one hand, some parameters have been defined as fuzzy values: the expectation [32], the distribution function in a point [5, 22], the variance [21] and the covariance[1] [27]. On the other hand, the expectation [24], the variance [13, 17, 25], the covariance[13] or the inequality index [26] have also been defined as crisp values. The introduction of these last definitions is guided by the interest of the authors in solving decision problems involving parameters with numerical, not fuzzy values.

In spite of the great amount of studies about fuzzy random variables, there are few works that study the different interpretations that could be given to their various definitions. The same can be said about the new concepts arising from them (for instance, some of the mentioned parameters.) It is well known that fuzzy sets admit of many different meanings (see, for example [12]) and each one of these meanings could lead to an interpretation of the concept of fuzzy random variable.

In this work, we shall observe that there are different extensions of the concept of variance to fuzzy random variables. We shall review different definitions of variance, found in the literature, and we shall propose an additional definition, that could be cast in a model of imprecise probabilities. We pay attention to the interpretation of each definition. Guided by simple examples, we shall observe the advantages and drawbacks of each definition in different contexts.

[1] We 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.

2 Fuzzy random variables

It was mentioned in the introduction that a fuzzy random variable is a function that assigns a fuzzy subset to each outcome of a random experiment. The different definitions of fuzzy random variable differ in the measurability conditions imposed to the random variables.

Kwakernaak [23] and Puri and Ralescu [32] rely on the α-cut mappings (the multivalued functions that map each element in the initial probability space to the respective α-cuts of the fuzzy set-valued image) to translate such condition. While Kwakernaak restrict himself to images that are fuzzy subsets of $\mathbb{R}$ and the boundaries of the α-cuts are measurable functions, Puri and Ralescu impose that the graph of the images itself be measurable (i.e. lies in the product σ-algebra.) On the other hand, Klement et al. [16] and Diamond and Kloeden [9] consider different metrics over the class of fuzzy sets of the output space and impose that the function is measurable with respect to the Borel σ-algebra induced by the corresponding metric. Krätschmer [18] reviews all the previous concepts and offers a unified vision when he considers certain topology, defined over the class of fuzzy subsets of $\mathbb{R}^n$, with non empty compact α-cuts. In this work, we shall not deal with formal aspects of each particular definition, but with the interpretation of the various concepts of fuzzy random variable.

Fuzzy sets have been given different interpretations [12], therefore a fuzzy random variable admits of various meanings as well. In the remaining part of this section, we briefly review two existing interpretations of fuzzy random variables, and introduce a new one. For every interpretation, we shall describe the information provided by the fuzzy random variable by means of a specific underlying model, namely, a classical probability model, an order 2 imprecise probability model and an order 1 imprecise probability model, respectively.

2.1 Linguistic random variables

In [32], Puri and Ralescu consider that the observations of some random experiments do not consist of numerical outputs, but are represented by vague linguistic terms. According to this idea, some authors consider that a fuzzy random variable is a measurable function, in the classical sense, between certain σ-algebra of events in the original space and a σ-algebra defined over a class of fuzzy subsets of $\mathbb{R}$. In this context, the probability distribution induced by the fuzzy random variable can be used to summarize the probabilistic information that the variable provides. If the fuzzy random variable has a finite number of images forming a linguistic term set, probability values can be assigned to the different linguistic labels. For example, the following model could be generated: the result is "high" with probability 0.5, "medium" with probability 0.25 and "low" with probability 0.25, where "high", "medium" and "low" are linguistic labels associated to fuzzy subsets of $\mathbb{R}$.

2.2 Ill-known classical random variables

On the contrary, Kruse and Meyer [22] 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, $X_0 : \Omega \to \mathbb{R}$, to which they refer to as the "original random variable." Therefore, the membership degree of a point x to the fuzzy set $\tilde{X}(\omega)$ represents the possibility degree of the assertion "$X_0(\omega)$ is x." (The image of element ω coincides with x.) This way, the authors get all the elements needed to define a possibility measure over the set of all random variables. They define the "acceptability degree" of each random variable, $X : \Omega \to \mathbb{R}$, as the value: $\mathrm{acc}(X) = \inf_{\omega\in\Omega} \tilde{X}(\omega)(X(\omega))$. The function "acc" take values in the unity interval. Therefore, it can be regarded as the possibility distribution associated to a possibility measure, $\Pi_{\tilde{X}}$, defined over the set of all random variables. $\mathrm{acc}(X)$ represents the possibility degree of X being the true random variable that models the studied experiment. If the fuzzy random variable was a random set (its images are crisp subsets of $\mathbb{R}$,) the acceptability function would assign the value 1 to random variables in a certain set, and the value 0 to the remaining ones. In the particular case when the fuzzy random variable is a classical random variable (all images are 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. In this case, its observation is completely precise.

Under this framework, we can build (see [6]) a possibility measure over the set of all the probability distributions in $\mathbb{R}$. The possibility distribution, $\pi_{P_{\tilde{X}}}$, that characterizes such possibility measure is defined as follows:

$$\pi_{P_{\tilde{X}}}(Q) = \sup\{\mathrm{acc}(X) \mid P_X = Q\} =$$

$$\Pi_{\tilde{X}}(\{X : \Omega \to \mathbb{R} \text{ measurable} \mid P_X = Q\}).$$

$\pi_{P_{\tilde{X}}}(Q)$ represents the degree of possibility that the original random variable is one of those that induce the probability distribution Q in $\mathbb{R}$). The possibility measure $\Pi_{P_{\tilde{X}}}$ is a "second-order possibility" formally

equivalent to those considered in [8]. It is so called, because it is a possibility distribution defined over a set of probability measures.

A possibility measure on a set represents the same information as a family of probability measures on this set (the family of probability measures that are dominated by the possibility measure and dominate the dual necessity measure[11].) Therefore, a second-order possibility measure is associated to a set of (meta-) probability measures, each of them defined, in turn, over a set of probability measures. Thus, a second-order possibility allows us to state assertions like "the probability that the true probability of the value 7 is 0.5 is between 0. and 0.7."

2.3 Known random process with imprecisely perceived output

Also in accordance with the possibilistic interpretation of fuzzy sets, in this work we are going to proceed in a slightly different way, in order to describe the information provided by $\tilde{X}$. We follow the path started in [31] for the particular case of the random sets and continued in [1] and [4] for fuzzy random variables. Suppose we have partial information about the probability distribution that models a sequence of two random experiments whose sample spaces are Ω and $I\!R$, respectively. For instance, the first one describes some random phenomenon of interest and the second one accounts for a measurement process applied to outcomes of the first one. Let us suppose, on the one hand, that the probability distribution that models the first one, $P : \mathcal{A} \to [0,1]$, 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 family of conditional possibility measures $\{\Pi(\cdot \mid \omega)\}_{\omega\in\Omega}$, each of them inducing the fuzzy set $\tilde{X}(\omega)$. This family of possibility measures models our knowledge about the relationship between the outcome of the first sub-experiment an the possible outcomes of the second one. (If the result of the first experiment is ω, then the possibility degree of x occurring in the second one is $\tilde{X}(\omega)(x)$.) In other words, we know the probability measure that drives the primary random process but the measurement process of outcomes is tainted with uncertainty.

The combination, using natural extension techniques [33] of both sources of information, allows to describe the available information about the probability distribution on $\beta_{I\!R}$ (the probability distribution that rules the second sub-experiment) by means of an upper probability (a standard imprecise probability model, not an order-2 model, like the one described before.) This way, we are able to state assertions like the following: "the probability of observing an outcome between 3 and 7 lies between 0.3 and 0.6."

Remark The mapping $X_0 : \Omega \to I\!R$ from last subsection represents a deterministic conditional probability $(P(\{X_0(\omega)\})|\omega) = 1, \forall\, \omega \in \Omega)$ only known as restricted by $\Pi(\cdot \mid \omega)$. Now the underlying conditional probability $P(\cdot|\omega)$ is not assumed to be deterministic.

3 Several definitions of variance

Each of the three models described in the preceding section leads a different understanding of the variance. In this section we consider the different definitions, according to each model, and emphasize their usefulness in different contexts. We shall restrict ourselves to the case where the images of the fuzzy random variable are fuzzy subsets of $I\!R$. In the last section of this work, we shall make some considerations about the generalization to the multi-dimensional case. Let $E(\tilde{X})$ be the expectation of the fuzzy random variable, in the sense of [32]. It is the fuzzy integral of the fuzzy random variable $\tilde{X}$: its α-cut is the Aumann integral of the set-valued mapping that to each ω assigns the set $\tilde{X}_\alpha(\omega) = \{r, \mu_{\tilde{X}(\omega)}(r) \geq \alpha\}$.

3.1 Classical model

Let us consider a probability space, $(\Omega, \mathcal{A}, P)$, and a metric, d, defined over the class of the fuzzy subsets of $I\!R$ (or over a subclass) and let us suppose that $\tilde{X} : \Omega \to \tilde{\mathcal{P}}(I\!R)$ is a function $\mathcal{A} - \beta(d)$-measurable (here, $\beta(d)$ represents the Borel σ-algebra induced by d.)

Definition 1 *We call* classical variance *of* $\tilde{X}$ *to the quantity*

$$\mathrm{Var}_{\mathrm{Cl}}(\tilde{X}) = \int_\Omega d(X, E(\tilde{X}))^2 \, dP.$$

The different definitions of variance in the literature that fit this formulation differ in the used metric and in the definition of the expectation of a fuzzy random variable. With respect to this, we briefly comment some details about the definitions of Körner ([17]) and Lubiano et al. ([25]). On the one hand, Körner considers Fréchet's expectation definition ([15]) for measurable functions taking values in a metric space. It is remarked that Fréchet defines the expectation of a measurable function Z, with values in a metric space (M, d) as a solution $E^{(d)}(Z)$, (not necessarily unique) of the problem $E[d(Z, E^{(d)}(Z))^2] = \min_{a\in M} E[d(Z, a))^2]$. Körner ([17]) checks that Puri and Ralescu's expectation ([32]) is the only Fréchet expectation for certain family of metrics defined over

the class of compact and normal fuzzy sets of $\mathbb{R}$, which they generically denote ρ_2. According to this, given a distance ρ_2, the variance of a fuzzy random variable $\tilde{X}$ is the amount

$$\mathrm{Var}_{\rho_2}(\tilde{X}) = \int_{\Omega} \rho_2(\tilde{X}, E_{PR}(\tilde{X}))^2 \, dP. \qquad (1)$$

With respect to the family of variances defined by Lubiano *et al.* in [25], the considered expectation $E_{PR}(\tilde{X})$ is also that of Puri-Ralescu, and the class of distances is also that defined by Bertoluzza et al. in [3], which in turn is a subclass of the family defined by Körner. In [17] and [25] we can find some interesting properties of the families of variances defined there. In this work we only comment some particular aspects of those, to show some advantages and also some drawbacks that they pose, if compared to other definitions of variance. Even though these definitions are stated for general fuzzy random variables in ([17]) and [25], in this work , it is sufficient to use their formulation in the particular case when $\tilde{X}$ is a multi-valued mapping (a function whose images are "crisp" subsets of the final space.)

In this case, we can easily check that the definitions of Körner and Lubiano et al. are of the form:

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

where $\pi_i \geq 0$, $i = 1, 2, 3$, $\pi_1 + \pi_2 + \pi_3 = 1$ 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.) Therefore, in the particular case in is a convex linear combination of the variances of their boundaries [2]. Rather, if $\pi_3 = (1 - \sqrt{\pi_1})^2$, the variance of $\tilde{X}$ coindices 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 image of the fuzzy random variable, and then calculate the variance of the classical random variable that results. The idea of computing the scalar variance using a representative substitute point to each fuzzy observation is used by Baudrit et al. [2], as one piece of information to be extracted from the hybrid propagation of fuzzy and probabilistic information through a mathematical model. With these examples, we observe that families of variances so defined allow us to quantify the dispersion of the (fuzzy, or set-valued) images of $\tilde{X}$, regarded as a measurable function from a classical point of view, and that can be useful when the images of $\tilde{X}$ are linguistic labels. In the context of a linguistic variable, $Var(\tilde{X})$ thus evaluates the variation across the possible linguistic labels.

The following example illustrates the shortcomings of this "classical" variance when quantifying the information available about the variance of an underlying random variable, when a "possibilistic" view of the fuzzy random variables is used, instead of the above setting.

Example 1 *Let us consider first a unitary sample space (that models a deterministic experiment,)* $\Omega_1 = \{\omega_1\}$, *with structure of probabilistic space with the only* σ*-algebra that can be defined over it,* $\mathcal{A}_1 = \mathcal{P}(\Omega_1)$ *and the only probability measure* P_1 *that is possible. Let us define a random set* $\Gamma_1 : \Omega_1 \to \mathcal{P}(\mathbb{R})$, *as* $\Gamma_1(\omega_1) = [-K, K]$. *In this example,* Γ_1 *serves us to represent the output (imprecisely known) of a deterministic experiment. For example, it allows us to represent the amount of money that, with absolute confidence, we shall receive, if we only know that it lies between* $-K$ *and* K. *Then, let us also consider another probabilistic space* $(\Omega_2, \mathcal{A}_2, P_2)$ *that corresponds to the outcome of tossing a fair coin (*$\Omega_2 = \{h, t\}$*), and the random set* $\Gamma_2 : \Omega_2 \to \mathcal{P}(\mathbb{R})$ *defined as* $\Gamma_2(h) = \Gamma_2(t) = [-K, K]$). *In turn,* Γ_2 *can be used to represent our gain after tossing a fair coin: the amount we are going to receive depends on the coin. The two outcomes are fixed before we perform the experiment, but we only know them in an imprecise manner, and actually we have the same knowledge* $[-K, K]$ *about these different values. The random sets* Γ_1 *and* Γ_2, *if regarded as classical measurable functions on the power set, induce the same possibility distribution (degenerated in the interval* $[-K, K]$*). Therefore, they both have the same Aumann expectation (that coincides with their own image) and they have null "classical" variance, since they are constant set-valued functions. But, if we follow Kruse and Meyer's, interpretation, we suppose that each of the maps models the imprecise observation of a classical random variable. Let us recall that, when the fuzzy random variable is reduced to multi-valued mapping,* Γ, *the information that it provides us about the original random variable* X_0, *can be interpreted as follows: for every* $\omega \in \Omega$, *all we know about the image of* ω, $X_0(\omega)$, *is that it is in the set* $\Gamma(\omega)$. *So, returning to the example, in the case of* Γ_1, *we are certain that the variance of the original random variable is 0. In the second case* (Γ_2) *we only know that it is a value between 0 and* K^2.

In practice, a fuzzy random variable can also be used to represent the imprecise observation of certain prop-

[2] The definition given by Feng in [13] and cited in the introduction fits this formulation for $\pi_1 = \pi_3 = 0.5$.

erty of the elements of a population Ω. To represent the information provided by the imprecise observations about the variance of the (classical) underlying random variable that models this property, we must resort to the variance defined by Kruse and Meyer.

3.2 Second-order imprecise model

In [21], Kruse defines the variance of a multi-valued mapping, $\Gamma : \Omega \to \mathcal{P}(\mathbb{R})$, as the set:

$$\mathrm{Var}_{\mathrm{Kr}}(\Gamma) = \{\mathrm{Var}(X) \mid X \in S(\Gamma)\},$$

where $S(\Gamma)$ represents the set of all measurable selections of the multi-valued mapping. The preceding definition can be easily extended to the case of fuzzy random variables as follows:

Definition 2 *Let us call* Kruse's variance *of the fuzzy random variable $\tilde{X} : \Omega \to \tilde{\mathcal{P}}(\mathbb{R})$, to the only fuzzy set determined by the nested family of sets:*

$$F(\alpha) := \mathrm{Var}_{\mathrm{Kr}}(\tilde{X}_\alpha), \forall\, \alpha,$$

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

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

$$\pi(x) = \sup\{\alpha \in (0,1] \mid x \in \mathrm{Var}_{\mathrm{Kr}}(\tilde{X}_\alpha)\},\ \forall\, x \in \mathbb{R}$$

Let us notice that:

$$\{x \mid \pi(x) > \alpha\} \subseteq F(\alpha) \subseteq \{x \mid \pi(x) \geq \alpha\}, \forall\, \alpha \in (0,1).$$

Hence, it is easy to see that the following equalities hold:

$$\pi(x) = \sup\{\mathrm{acc}(X) \mid \mathrm{Var}(X) = x\},\ \forall\, x \in \mathbb{R}.$$

From now on, we shall denote by $\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})$ the fuzzy set with membership function π. It is clear that this definition is compatible with the second-order possibility model shown in Section 2. Therefore, the membership degree of a value x to the fuzzy set $\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})$ represents the maximal possibility degree of the original random variable among those whose variance is equal to x. See [20] for the computation of the empirical set-valued variance of a finite set of set-valued realizations and [10] for the fuzzy case.

When the outputs of a random experiment are imprecisely observed, our knowledge about their dispersion is also imprecise. So, Kruse's variance can be called *potential variance*, since $\pi(x)$ is the degree of possibility that x is the variance (in case it exists) of the actual underlying random variable. $\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})$ reflects the imprecision pervading the observation of the outcome of a random experiment. Therefore, it produces a crisp set of potentially attainable variances (when the imprecise observations of the random variable are set-valued) or a fuzzy set (when it is represented by a fuzzy random variable). It does not produce a real value, like the "classical" observable variance of the previous section. Thus, when the random set (or the fuzzy random variable) represents the imprecise observation of a "classical" 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. Let us show an illustrative example.

Example 2

(a) The set $\Omega = \{\omega_1, \ldots, \omega_4\}$ 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 $\Gamma(\omega_i) = [9.5, 10.5], \forall\, i = 1, \ldots, 4$. 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's variance gives us. Misleadingly, the classical variance of Γ returns the value 0.

(b) Case (a) is an example where the classical variance of the random set Γ is not an upper bound of the actual value of the variance of X_0. Neither it is, in general, a lower bound, as we are going to show. Let us suppose that four objects $\omega_1, \ldots, \omega_4$ 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 $\tilde{X}_0$ is given by the random set $\Gamma : \omega \to \mathcal{P}(\mathbb{R})$ 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, but the "classical" variance assigns a strictly positive value to it. On the other hand, Kruse's variance produces the interval $[0, 0.092]$.

The last case suggests that the observed classical variance of a fuzzy random variable can be misleading. It may reflect the variance *of the imprecision* of the output (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.

On the other hand, neither Kruse's variance is determined by the classical one, nor the converse holds. Let

us illustrate these ideas with the aid of the following examples.

Example 3 *Let us consider now the random sets in Example 1. According to Kruse, their respective variances represent the sets of possible values of the variance of the corresponding original random variable. Thus, in that example, the respective variances are, according to this definition,* $\mathrm{Var}_{\mathrm{Kr}}(\Gamma_1) = \{0\}$ *and* $\mathrm{Var}_{\mathrm{Kr}}(\Gamma_2) = [0, K^2]$. *However, the classical variance assigns the value 0 to both random sets.*

We observe that Kruse's variance allows us to distinguish between two fuzzy random variables with the same "classical" probability distribution when they are used in this context. However, it does not always associates different values to two fuzzy random variables with different "classical" variance, as we shall see in the example that follows.

Example 4 *Let us consider the probability space* $(\Omega_2, \mathcal{A}_2, P_2)$ *of the example 1 and the constant random set* $\Gamma_2 : \Omega_2 \to \mathcal{P}(I\!R)$, *defined there. Let us also define the random set* $\Gamma_3 : \Omega_2 \to \mathcal{P}(I\!R)$ *as follows:* $\Gamma_3(h) = [-K, 0]$ *and* $\Gamma_3(t) = [0, K]$. *In both cases, Kruse's variance produces the interval* $[0, K^2]$. *But the classical variance would assign the value 0 to* Γ_2 *and a strictly positive value to* Γ_3.

The last example serves us to observe that Kruse's variance does not allow, generally speaking, to quantify the dispersion of the images of a fuzzy random variable, when it is considered as a classical measurable function.

In fact, the scalar variance of section 3.1 could be used in the context of an imprecisely observed random variable, but it could only account for an "observable variance", namely the part of the variance that can be measured, despite the imprecision of the observation. Indeed in the example 1, the fair die case leads to a zero observable variance, because the variability of the die is drowned into the imprecision of the observation. However, it seems that the scalar variance, when non-zero, may partially account for the variability of the underlying phenomenon: 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 nondeterministic 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. Nor does a positive scalar variance 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. In fact, one way of computing the observable variance as a scalar is to choose an appropriate distance between fuzzy sets instead of ρ_2 in the scalar variance (1), namely one that vanishes when the two fuzzy intervals overlap: consider two fuzzy intervals F and G, and let

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

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 that the lower bound of Kruse's variance. In example 2(b), the above scalar variance is now 0, and so is it in example 4. In example 2(b), the Körner scalar variance essentially reflects the variability of the precision of the observation.

3.3 First-order imprecise model

In this section, we propose a model that also takes imprecision into account, although in a different manner. We consider here a first-order imprecise probabilities model, instead of a second-order one. Therefore, the new variance assigns a crisp set to every fuzzy random variable. With the help of easy examples, we shall show the similarities and differences between this new model and the present one.

The present definition of variance is based upon the first-order, imprecise probabilities model that was shown at the end of section 2. As we pointed out there, we consider, on the one hand, the probability measure P (defined over $\mathcal{A}$), that models a first subexperiment, and, on the other hand, a family of conditional possibility measures, $\{\Pi(\cdot \mid \omega)\}_{\omega\in\Omega}$, defined as follows:

$$\Pi(A \mid \omega) = \Pi_{\tilde{X}(\omega)}(A) = \sup_{x\in A} \tilde{X}(\omega)(x),\ \forall\, A \in \beta_{I\!R},\ \forall\, \omega.$$

In the preceding formula, $\Pi_{\tilde{X}(\omega)}$ represents the possibility measure determined by the possibilty distribution $\tilde{X}(\omega) : I\!R \to [0, 1]$. So, the value $\Pi(A|\omega)$ is an upper bound for the probability that the final outcome is in A, verifying the hypothesis that the outcome of the initial experiment is ω. This family of possibility measures represents our (imprecise) knowledge carried by $\tilde{X}$ about the relation that exists between the outcome of the first sub-experiment and the set of all the possible outcomes of the second one.

Therefore, the relationship between the two experiments is given by a **transition probability** $Q(\cdot|\cdot)$ on $\beta_{I\!R} \times \Omega$, i.e., a function such that:

1. $Q(\cdot|\omega)$ is a probability measure for all $\omega \in \Omega$.
2. $Q(A|\cdot)$ is $\mathcal{A} - \beta_{[0,1]}$-measurable for all $A \in \beta_{I\!R}$,

and the available knowledge about this transition probability is modelled by the conditional possibility measures $\{\Pi(\cdot|\omega)\}_{\omega\in\Omega}$, in the sense that $Q(\cdot|\omega) \leq \Pi(\cdot|\omega)$ for all $\omega \in \Omega$.

Within this context, all we know about the probability distribution that models the second experiment is that it is in the set $\mathcal{C}_2 =$

$$\{Q_2 \mid Q_2 \text{ marginal of } P \text{ and } Q(\cdot|\cdot),\ Q(\cdot|\cdot) \in \mathcal{C}\}, \quad (2)$$

where

$$\mathcal{C} = \{Q(\cdot|\cdot) \ \mid Q(A|\omega) \leq \Pi(A|\omega)\ \forall\, A \in \beta_{I\!R},\ \omega \in \Omega\}.$$

It is easily observed that this is a generalization of the concept of probability induced by a classical random variable. Let us suppose that the images of the fuzzy random variable $\tilde{X}$ are real values. In other words, let us suppose that for all $\omega \in \Omega$, $\Pi(\cdot|\omega)$ is, in particular, the degenerated probability measure in a point $X(\omega)$. In this case, we are admitting a complete confidence about the relationship between both sub-experiments (if the result of the first sub-experiment is ω, then we are absolutely certain the outcome of the second experiment is $X(\omega)$). It is easy to prove that the class $\mathcal{C}_2$ in equation (2) is reduced to the singleton $\{P_X\}$ (in this case, the probability induced by $X : \Omega \to I\!R$ in $\beta_{I\!R}$ is the only probability measure compatible with P and $\Pi(\cdot|\cdot)$). Besides, the variance of a classical random variable, $\mathrm{Var}(X) = \int_\Omega [X - E(X)]^2 dP$, can be alternatively expressed as the following Lebesgue integral with respect to P_X:

$$\mathrm{Var}(P_X) = \int_{I\!R} \left(\mathrm{id} - \int_{I\!R} \mathrm{id}\, dP_X\right)^2 dP_X,$$

where id:$I\!R \to I\!R$ is the identity function[3]. Therefore, in the proposed imprecise probabilities model proposed, all we know about the variance of the output of the second sub-experiment is that it belongs to the set $\mathrm{Var}_{\mathrm{Im}\ 1}(\tilde{X})$ defined as follows:

Definition 3 *Consider a probability space $(\Omega, \mathcal{A}, P)$, and a fuzzy random variable defined over it, $\tilde{X} : \Omega \to \tilde{\mathcal{P}}(I\!R)$. 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:*

$$\mathrm{Var}_{\mathrm{Im}-1}(\tilde{X}) = \{\mathrm{Var}(Q_2) \mid Q_2 \in \mathcal{C}_2\}$$

where

$$\mathcal{C}_2 = \{Q_2 \mid Q_2 \text{ marginal of } P \text{ and } Q(\cdot|\cdot),\ Q(\cdot|\cdot) \in \mathcal{C}\},$$

and $\mathcal{C} = \{Q(\cdot|\cdot)$ transition prob. $\mid$
$Q(A|\omega) \leq \Pi(A|\omega)\ \forall\, A \in \beta_{I\!R},\ \omega \in \Omega\}$.

$\mathrm{Var}_{\mathrm{Im}-1}(\tilde{X})$ is the set of possible values of the variance of the second sub-experiment, according to the available information. We are going to compare, in an example, the information provided by $\mathrm{Var}_{\mathrm{Im}\,1}$ and $\mathrm{Var}_{\mathrm{Kr}}$ about the variance of the "original" probability distribution.

Example 5 *Let us consider the unit interval, $\Omega = [0,1]$, equipped with the Lebesgue measure. Let us also consider the fuzzy random variable $\tilde{X} : \Omega \to \tilde{\mathcal{P}}(I\!R)$ constant in the fuzzy set $\tilde{A}$ determined by the α-cuts $\tilde{A}_\alpha = [-(1-\alpha), 1-\alpha]$. It can be easily checked that $[\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})]_\alpha = [0, (1-\alpha)^2]$, $\forall\, \alpha > 0$. On the other hand, we can observe that $\mathrm{Var}_{\mathrm{Im}-1}(\tilde{X})$ is the interval [0,1/3][4]. It is clear that this interval is strictly contained in the support of $\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})$. Therefore, under the first-order model here described, the variance of the results of the experiment is known to be less or equal that 1/3, while under the second-order probability model, a strictly positive possibility degree is also assigned to all variables between 1/3 and 1.*

Despite the fact that the two models considered in last example (orders 1 and 2 imprecise probability models) are associated to a possibilistic interpretation of fuzzy sets, the meaning of the two definitions of variance derived from them are quite different. In the second-order model, the fuzzy random variable, $\tilde{X}$, represents an imprecise observation of a particular (classical) random variable, $X_0 : \Omega \to I\!R$. 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 first-order model, the fuzzy random variable $\tilde{X}$ represents our (imprecise) knowledge about the link between two steps of a random experiment. Thus, the same result ω in the first step can be associated to different outcomes of the second step. Under the first-order model 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 this last probability measure is given by a pair of upper-lower probability measures, so is our knowledge about the probability measure that governs the second step.

Let us examine now the relation between both models in the particular case where $\tilde{X}$ is a random set. ($\tilde{X}(\omega)$ is a crisp set, $\forall\, \omega \in \Omega$.) In this case, Kruse's variance is defined as:

$$\mathrm{Var}_{\mathrm{Kr}}(\tilde{X}) = \{\mathrm{Var}(P_X) \mid X \in S(\tilde{X})\} =$$

[3]Since the variance of a classical random variable is a function of its induced probability distribution,we shall commit a small abuse of the language from now on and we shall express it as the variance of such probability distribution.

[4]It is actually equal to $\frac{1}{2}\int_0^1 (\inf A_\alpha - \sup A_\alpha)^2 d\alpha$. See Dubois et al.[10].

$$\{\mathrm{Var}(Q) \mid Q \in \mathcal{P}(\tilde{X})\},$$

where $\mathcal{P}(\tilde{X})$ is the set of probability measures associated to the measurable selections of $\tilde{X}$,

$$\mathcal{P}(\tilde{X}) = \{P_X | X \in S(\tilde{X})\}.$$

On the other hand, the first-order imprecise variance is given by the formula:

$$\mathrm{Var}_{\mathrm{Im}\,1}(\tilde{X}) = \{\mathrm{Var}(Q_2) \mid Q_2 \in \mathcal{C}_2\}, \text{ where}$$

$$\mathcal{C}_2 = \{Q_2 \mid Q_2 \text{ marginal of } P \text{ and } Q(\cdot|\cdot), Q(\cdot|\cdot) \in \mathcal{C}\},$$

and $\mathcal{C}$ is the set of transition probability measures:

$$\mathcal{C} = \{Q(\cdot|\cdot) \mid Q(A|\omega) \leq \Pi(A|\omega)\ \forall\, A \in \beta_{I\!R},\ \omega \in \Omega\}.$$

In the above formula, $\Pi(\cdot|\omega)$ is the Boolean possibility measure associated to the (crisp) set $\tilde{X}(\omega)$. For an arbitrary measurable selection of $\tilde{X}$, $X \in S(\tilde{X})$, and a fixed $\omega \in \Omega$, let us consider the probability measure degenerated on the point $X(\omega)$, $\delta_{X(\omega)}$. Let us construct the function $Q(\cdot|\cdot) : \beta_{I\!R} \times \Omega \to [0,1]$ as $Q(\cdot|\omega) = \delta_{X(\omega)}$, $\forall\, \omega \in \Omega$. It is easy to see that $Q(\cdot|\cdot)$ is a transition probability measure and it belongs to the set $\mathcal{C}$. So the probability measure $P_X : \beta_{I\!R} \to [0,1]$ belongs to $\mathcal{C}_2$. Thus, we observe that the set $\mathcal{P}(\tilde{X})$ is included in $\mathcal{C}_2$ and so $\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})$ is contained in $\mathrm{Var}_{\mathrm{Im}\,1}(\tilde{X})$. Furthermore $\mathcal{C}_2$ is a convex set of probability measures, but $\mathcal{P}(\tilde{X})$ is not convex in general. (The properties of $\mathcal{P}(\tilde{X})$ are studied in detail in [6, 7, 28, 29, 30].) These differences can influence the calculations of the variances, as shown in the following example.

Example 6 *Consider again the random sets used in example 1. According to the model described in that section, in the first case the first sub-experiment is deterministic, and the relationship between both sub-experiments is determined by* Γ_1. *This random set represents an "empty" conditional probability distribution over* $[-K, K]$. *Therefore, the set of conditional probability measures* $Q(\cdot|\omega_1)$, *that are compatible with them is the set of all measures that assign probability 1 to the set* $[-K, K]$. *This way, the following information is given: once the first experiment is performed, a random number between* $-K$ *and* K *is chosen, and not a number selected beforehand. This is the difference between the second-order model described before and the current model. In the second-order model, the number was selected beforehand, but it was unknown. Now, in the case of* Γ_2, *the first sub-experiment consists in tossing a coin. Once the result has been observed, it is chosen, whatever the result is, a random number between* $-K$ *and* K. *Therefore, in this example is intuitively clear that, regarding the outcome of the second sub-experiment, we could obviate tossing the coin (we could not in the second order model) and then* Γ_1 *and* Γ_2 *show, according to the interpretation of the first-order model the same information. Thus we observe that* $\mathrm{Var}_{\mathrm{Im}-1}(\Gamma_1) = \mathrm{Var}_{\mathrm{Im}-1}(\Gamma_2) = [0, K^2]$.

Let us comment on some relationships that exist between the variance of this imprecise, first-order model, and the classical variance of section 3.1.

We easily observe that none of them can be calculated as a function of the other one:

Example 7 *Let us consider, on the one hand, the random set* Γ_1 *defined in the example 1 and, on the other, the random set* Γ_5, *defines over the same space, of the form* $\Gamma_5(\omega_1) = \{0\}$. *The "classical" variance assigns the value 0 to both random sets, while the imprecise variance assigns the set of values* $[0, K^2]$ *to the first problem, and the singleton* $\{0\}$ *to the second.*

In a similar manner, we can check that the classical variance can not either be expressed as a function of the variance that is considered in this section. It is enough to observe the random sets of the example 4.

4 Concluding remarks

In this work we have studied different proposals to generalize the concept of variance of a real random variable to fuzzy random variables. In Körner 's work ([17]) is stated a more general definition, valid when the final space is $I\!R^n$, with arbitrary $n \in I\!N$. In that work, the variance of $\tilde{X}$ is defined as the expectation of the squares of the distances of their images to their Fréchet expectation. 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 images of the fuzzy random variable, but not directly related to the concept of variance-covariance matrix.

If, on the contrary, the aim is to generalize the latter concept, Kruse's procedure can be applied without too many changes. Using similar reasoning methods as those of this author, it can be obtained a fuzzy set over the class of square matrices, that associates, to each particular matrix, a degree of possibility. This fuzzy set models the imprecise knowledge available about the variance-covariance matrix of the "original" random vector. In [27], Meyer proposes a definition of covariance following a path similar to Kruse's. According to our intuition, the combination of the information provided separately about the variance of

every component and about the covariance between them is more imprecise that the straight information about the variance-covariance matrix.

With respect to the different definitions of variance considered in this work, we think that none of them is, in general terms, preferable to the others, but they either serve different purposes or reflect different models of the observed phenomenon, as well as different observation settings. Therefore, it should be decided whether the dispersion needs to be measured as a number, a fuzzy set or a crisp set. If the fuzzy random variable is interpreted as a classical measurable function, the most appropriate decision would involve using Feng, Körner or Lubiano et al.'s definitions. It measures the variability of the observed membership function, not the variability of the quantity it possibly describes. Such classical definitions do not take into account any kind of imprecision, but they merely quantify the dispersion of the (fuzzy) images of the fuzzy random variable. Some of these classical definitions are equivalent to considering first a representative (numerical) element of every image of the fuzzy random variable (the center point of the 0-cut, for instance) and then calculate the dispersion of these numerical values.

Part of the actual variability can be observed and measured by means of a scalar if the fuzzy outcomes are precise enough and often disjoint. On the other hand, the average precision of the fuzzy random variable, and the variance of the precision are other useful evaluations.

If the fuzzy random variable represents an imprecise measurement of certain characteristic of the elements of the sample space, one of the two non-scalar definitions must be used. For example: let us suppose we intend to calculate the dispersion of the weights of a bunch of apples, and we use an imprecise scale. Let us suppose that, for every confidence level $1-\alpha$ we know that the real weight is at most at d_α from the value produced by the scale. In this case, every α-cut of Kruse's variance represents our knowledge about the true dispersion of the weights of the apples, for every confidence level $1-\alpha$. On the other hand, the variance proposed in section 3.3 represents the set of all possible values for the dispersion of the weights, if we combine the initial randomness (tied to the random experiment "choose an apple") with the randomness originated in the degrees of confidence associated to the scale accuracy. Therefore, if the fuzzy random variable represents the knowledge about the relationship between the two sub-experiments ("if we choose the apple ω, the degree of possibility of its weight is x is $\tilde{X}(\omega)(x)$"), then the definition proposed in section 3.3 should be used.

Acknowledgements

This work has been supported by grants MTM2004-01269 and TIC2002-04036-C05-05. Both grants partially participate to FEDER funds.

References

[1] C. Baudrit, I. Couso, D. Dubois (2005) Probabilities of events induced by fuzzy random variables, *Proceedings of the Conference EUSFLAT-LFA 2005*. Barcelona.

[2] C. Baudrit, D. Dubois, D. Guyonnet, H. Fargier. Joint Treatment of Imprecision and Randomness in Uncertainty Propagation. In : Modern Information Processing: From Theory to Applications. B. Bouchon-Meunier, G. Coletti, R.R. Yager (Eds.), Elsevier, p. 37-47, 2006.

[3] C. Bertoluzza, A. Salas, N. Corral (1995) On a new class of distances between fuzzy numbers. *Mathware and Soft Computing* **2** 71-84.

[4] 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. Hyrniewicz, and J. Lawry), 133-140. Springer, Heidelberg.

[5] I. Couso, S. Montes, P. Gil (1998) Función de distribución y mediana de variables aleatorias difusas, *Proceedings of the Conference ESTYLF'98*. Pamplona. (In Spanish)

[6] 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.

[7] I. Couso, L. Sánchez, P. Gil (2004) Imprecise distribution function associated to a random set, *Information Sciences* **159** 109-123.

[8] G. de Cooman, P. Walley (2002) An imprecise hierarchical model for behaviour under uncertainty, *Theory and Decision* **52** 327-374.

[9] P. Diamond, P. Kloeden (1994) *Metric Spaces of Fuzzy Sets*, World Scientific, Singapur.

[10] D. Dubois, H. Fargier, J. Fortin. The empirical variance of a set of fuzzy intervals. Proc. of the 2005 IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE 2005), Reno, Nevada, 22 -25 mai 2005. IEEE Press, p. 885-890.

[11] D. Dubois, H. Prade (1987) The mean value of a fuzzy number, *Fuzzy Sets and Systems* **24** 279-300.

[12] D. Dubois, H. Prade (1997) The three semantics of fuzzy sets, *Fuzzy Sets and Systems* **90** 141-150.

[13] 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.

[14] R. Féron (1976) Ensembles aléatoires flous, *C.R. Acad. Sci. Paris Ser. A* **282** 903-906.

[15] M. Fréchet (1948) Les éléments aleatoires de natures quelconque dans un espace distancié, *Ann. Inst. H. Poincaré* **10** 215-310.

[16] E.P. Klement, M.L. Puri, D.A. Ralescu (1986) Limit theorems for fuzzy random variables *Proc. Roy. Soc. London A* **407** 171-182.

[17] R. Körner (1997) On the variance of fuzzy random variables, *Fuzzy Sets and Systems* **92** 83-93.

[18] V. Krätschmer (2001) A unified approach to fuzzy random variables, *Fuzzy Sets and Systems* **123** 1-9.

[19] V. Krätschmer (2003) When fuzzy measures are upper envelopes of probability measures, *Fuzzy Sets and Systems* **138** 455-468.

[20] 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.

[21] R. Kruse (1987) On the variance of random sets, *J. Math. Anal. Appl.* **122** 469-473.

[22] R. Kruse, K.D. Meyer (1987) *Statistics with vague data* D. Reidel Publishing Company.

[23] Kwakernaak (1989) Fuzzy random variables. Definition and theorems. *Inform. Sci.* **15** 1-29.

[24] Y.K. Liu, B. Liu (2003) A class of fuzzy random optimization: expected valued models, *Information Sciences* **155** 89-102.

[25] M.A. Lubiano, M.A. Gil, M. López-Díaz, M.T. López-García (2000), The $\vec{\lambda}$-mean squared dispersion associated with a fuzzy random variable, *Fuzzy Sets and Systems* **111** 307-317.

[26] 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.

[27] 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.

[28] E. Miranda, I. Couso, P. Gil (2002) Upper probabilities and selectors of random sets. *En: Soft methods in probability, statistics and data analysis* P. Gzegorzewski, O. Hryniewicz, M.A. Gil (Eds.). Physica-Verlag, Heidelberg, Alemania.

[29] E. Miranda, I. Couso, P. Gil (2003) Study of the probabilistic information of a random set, *Proceedings of the 3rd ISIPTA Conference*. Lugano, Suiza.

[30] E. Miranda, I. Couso, P. Gil (2005) Random sets as imprecise random variables. *Journal of Mathematical Analysis and Applications* **307** 32-47.

[31] E. Miranda, G. de Cooman, I. Couso (2005) Imprecise probabilities induced by multi-valued mappings, *J. Stat. Plann. Inference.* **133** 173-197.

[32] M.L. Puri, D. Ralescu (1986) Fuzzy Random Variables, *J. Math. Anal. Appl.* **114** 409-422.

[33] P. Walley (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.