

Specific Features of Descriptive Statistics with Fuzzy Random Variables

Oļegs Užga-Rebrovs¹, Gaļina Kulešova²

¹ Rezekne Academy of Technologies, Rezekne, Latvia,

² Riga Technical University, Riga, Latvia

Abstract – The aim of the paper is to examine procedures of descriptive statistics in the case when the values of relevant attribute in a sample are set in the form of fuzzy categories. The paper provides alternative definitions of a fuzzy random variable, and describes corresponding procedures for calculating the analogues of location and spread parameters. The paper also presents some illustrative examples and analyses the results obtained. Based on the result analysis, practical recommendations are given on how to use procedures of fuzzy statistics.

Keywords – Deterministic variance, fuzzy mean value, fuzzy random variable in Kwakernaak's sense, fuzzy random variable in Puri-Ralescu's sense, fuzzy variance.

I. INTRODUCTION

Classical statistics is a powerful tool for processing a different kind of initial data. Based on the descriptive statistical procedures, it is possible to calculate parameters of location and spread for sample initial data distribution. By means of procedures of inference statistics, the obtained results can be transferred to the population, and potential errors related to such transfer can be evaluated.

One limitation of classical statistics is that the values of relevant attribute in the sample and/or population have to be set in the form of real numbers. However, in many practical situations, it is impossible to get them. Relevant evaluations can only be obtained in uncertain form as intervals or fuzzy categories.

II. ALTERNATIVE DEFINITIONS OF A FUZZY RANDOM VARIABLE

This section introduces two alternative conceptual definitions of the notion of a fuzzy random variable. These definitions are essential for further analysis and conclusions.

Let us assume that a sample of individuals is available, whose attribute is their age. Initial data are schematically shown in Fig. 1.

Let us also assume that due to some reasons, it is impossible to evaluate real values of individual's age; and these values are evaluated using fuzzy categories, whose membership functions are shown in Fig. 2.

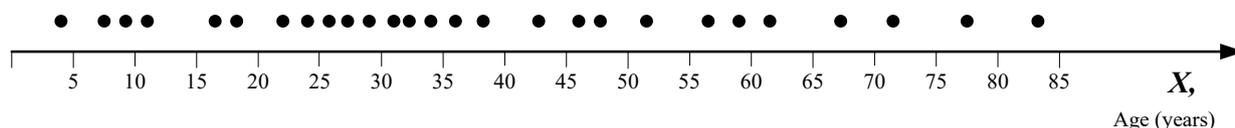


Fig. 1. Schematic representation of individuals' age in the sample.

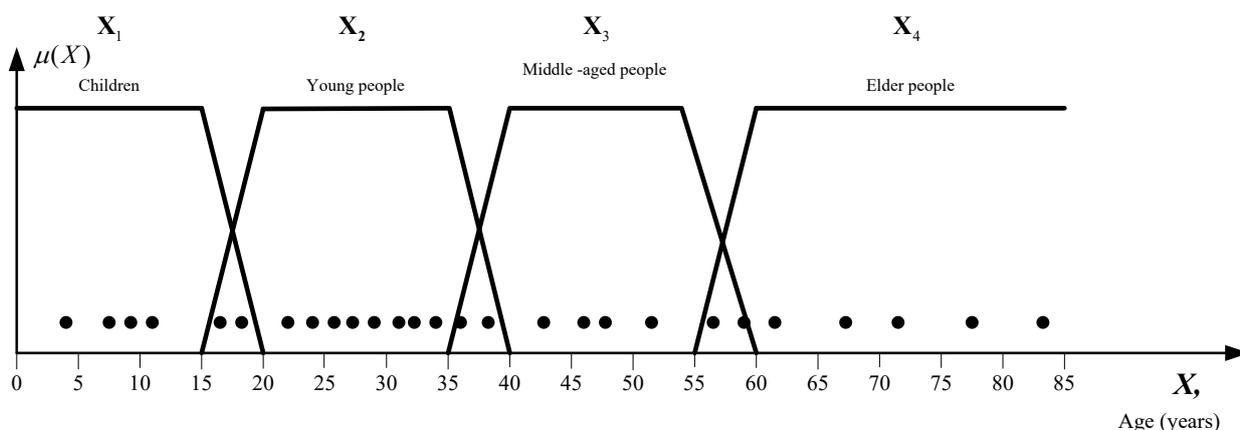


Fig. 2. Conditional graphs of membership functions for fuzzy categories of age.

For illustration purposes, this figure depicts real values of individuals age in the sample. But now, the initial data are represented as follows: individual x based on his/her age belongs to a fuzzy category $\mathbf{X}_i, i = 1, 2, 3, 4$.

A variable whose elements are expressed in the form of fuzzy categories is called a *fuzzy random variable*. Note that if the sense of fuzzy categories is expressed by means of linguistic terms, one can speak about linguistic categories [1].

Let us introduce formal concepts and definitions related to fuzzy random variables. Historically, the first was a definition given in [2], [3]. Let us cite the original definition from [2]. In that definition, (Ω, A, P) is a standard definition of the probabilistic space, where Ω is a set of outcomes, A is σ -algebra in the set of outcomes and P is a set of probabilistic values.

The notion of a fuzzy random variable will be introduced as follows. Let (Ω, A, P) be a probability triple. Suppose that U is a random variable defined on this triple. Assume now that we perceive this random variable through a set of windows $W_i, i \in J$, with J – a finite or countable set, each representing an interval of the real line, such that $W_i \cap W_j = \emptyset$ for $i \neq j$, and $\cup_{i \in J} W_i = R$. Perceiving the random variable through these windows means that for each ω we can only establish whether $U(\omega) \in W_i$ for some $i \in J$.

Let us define the function $I_i : R \rightarrow [0, 1]$ as the characteristic function of the set W_i . Let S be also the space of all piecewise continuous function mapping $R \rightarrow [0, 1]$. We then define the perception of the random variable U , as described above, as the mapping $X : \Omega \rightarrow S$ given by

$$\omega \rightarrow^X X_\omega$$

with $X_\omega = I_i$, if and only if $U(\omega) \in W_i$. This means that we associate with each $\omega \in \Omega$ not a real number $U(\omega)$, as in the case of an ordinary random variable, but a characteristic function X_ω , which is an element of S .

The map $X : \Omega \rightarrow S$ described above characterises a special type of fuzzy random variable. The random variable U , of which this fuzzy random variable is a perception, is called *original* of the fuzzy random variable. We note that there may be many originals that correspond to a given fuzzy random variable.

At this point, we generalize and define a fuzzy random variable as a map $\xi : \Omega \rightarrow F$, where F is a set of all fuzzy numbers. Let us denote the image of ω in F under ξ as $\xi(\omega) = (R, X_\omega, a_\omega)$, with $X_\omega \in S$ and $a_\omega : R \rightarrow P$. The map $X : \Omega \rightarrow S$, specified by

$$\omega \rightarrow^X X_\omega,$$

is required to be such that for each $\mu \in [0, 1]$ both U_μ^* and U_μ^{**}

$$U_\mu^*(\omega) = \inf \{x \in R / X_\omega(x) \geq \mu\},$$

$$U_\mu^{**}(\omega) = \sup \{x \in R / X_\omega(x) \geq \mu\},$$

are finite real-valued random variables defined on (Ω, A, P) satisfying

$$\forall \omega \in \Omega, X_\omega(U_\mu^*(\omega)) \geq \mu, X_\omega(U_\mu^{**}(\omega)) \geq \mu.$$

Finally, for each $\omega \in \Omega$ and each $x \in R$ $a_\omega(x)$ there is the statement $a_\omega(x) =$ (the original assumes the value x at the point ω), where we refer to the original random variable of which ξ is a fuzzy perception.

Kruse and Meyer [4], [5] have supplemented and defined Kwakernaak's definition more precisely; that is why, in the literature, the aforementioned fuzzy random variable is sometimes called fuzzy random variable in Kwakernaak-Kruse-Meyer's sense. In the paper, we will call such a variable fuzzy random variable in Kwakernaak's sense.

In the context of fuzzy random variables, mapping $\mathbf{X} : \Omega \rightarrow F_c(R)$ means that the set of outcomes is mapped to the set of fuzzy categories. For example, a set of individuals is mapped into a set of fuzzy categories corresponding to their age.

Another conceptual interpretation of fuzzy random variable is given in [6]. Let us cite the corresponding definition from [7].

Let (Ω, A, P) be a probability space modelling a random experiment. Let $F(R^p)$ be a class of fuzzy subsets $\tilde{U} \rightarrow [0, 1]$, such that $\tilde{U}_\alpha = \{x \in R^p : \tilde{U}(x) > \alpha\}$ is compact for all $\alpha \in [0, 1]$ and $\tilde{U} \neq \emptyset$.

Fuzzy random variable in Puri-Ralescu's sense associated with (Ω, A, P) is a map $\mathbf{X} : \Omega \rightarrow F(R^p)$ such that for each value $\alpha \in (0, 1]$, fuzzy-valued map $X_\alpha \subset X_\alpha(\omega) = (X(\omega))_\alpha$ for all $\omega \in \Omega$ is a random compact set, i.e., Borel-measurable map regarding Borel σ -field generated by means of a topology associated with Hausdorff metric in the space of non-empty compact subsets of R^p [3], [7].

Unlike Kwakernaak's interpretation, Puri-Ralescu's interpretation does not provide any real scale, on which fuzzy categories of relevant attribute would be defined. Moreover, these values can be defined in different spaces of real numbers. The map $\mathbf{X} : \Omega \rightarrow F(R^p)$ explicitly demonstrates it. Besides, it is not assumed that original random variable exists, on whose values the values of the attribute are defined. Instead, attribute value evaluation is made directly in terms of fuzzy categories.

III. CALCULATION OF STATISTICAL PARAMETERS OF A FUZZY RANDOM VARIABLE

As pointed out in [8], probabilistic and statistical studies of fuzzy random variables are commonly connected to fuzzy-valued and real-valued "parameters". In essence, those parameters do not have the same sense as the one that takes place in the case of real-valued initial data. In the case of the fuzzy-valued initial data, they can be referred to as "aggregate evaluations" of fuzzy random sets that are components of a fuzzy random variable.

Taking into account that elements of a fuzzy random variable are fuzzy (linguistic) categories, special calculation expressions are needed to calculate its statistical parameters. For that purpose, it is necessary to introduce a suitable metric in the space of fuzzy numbers. The most proper metric in the space of fuzzy initial data would be the metric described in [9]. Let us briefly outline the metric based on the data provided in [10].

Rational point behind the distance between fuzzy numbers is, on the one hand, that it extends Euclidean space between the real-valued values and, on the other hand, it extends the suitable distance between intervals at α -levels of fuzzy numbers.

Based on the above requirements, a metric in [9] is defined at these assumptions:

- W is a normalizing weighted value in the measurable space $([0,1]B_{[0,1]})$, which is formalised as the probabilistic value related to the non-degenerate distribution,
- φ is another normalizing probabilistic value in the measurable space $([0,1]B_{[0,1]})$, which is formalised as the probabilistic value related to absolute continuous distribution function that is increasing exactly on $(0,1]$.

Using the introduced concepts, we can determine the (W, φ) distance as follows [10]:

$$D_W^\varphi(\tilde{U}, \tilde{V}) = \sqrt{\int_{(0,1]} [dW(\tilde{U}_\alpha, \tilde{V}_\alpha)]^2 d\varphi(\alpha)}, \quad (1)$$

where

$$dW(\tilde{U}_\alpha, \tilde{V}_\alpha) = \sqrt{\int_{(0,1]} [\tilde{U}_\alpha^{[t]} - \tilde{V}_\alpha^{[t]}] dW(t)}, \quad (2)$$

and

$$\tilde{U}_\alpha^{[t]} = t * \sup \tilde{U}_\alpha + (1-t) * \inf \tilde{U}_\alpha.$$

If fuzzy random variable \mathbf{X} is defined in Kwakernaak's sense, its distribution parameters are treated as extensions of the corresponding parameters of the original underlying real-valued random variable. Then fuzzy perception of the mean value $\tilde{E}(\mathbf{X})$ in common case can be formally represented as follows:

$$\tilde{E}(\mathbf{X})(t) = \sup_{X_0 \in \text{Orig}(\mathbf{X}), E(X_0)=t} \inf_{\omega \in \Omega} \mathbf{X}(\omega)(X_0(\omega)) \quad (3)$$

This is valid for $\forall \alpha \in [0,1]$

$$E(\mathbf{X})_\alpha = [E(\inf X_\alpha), E(\sup X_\alpha)] \quad (4)$$

If fuzzy variable \mathbf{X} is defined in Puri-Ralescu's sense, its fuzzy mean value can be expressed in common case as follows:

$$\tilde{E}(\mathbf{X}) = \arg \min_{X_i \in F_c(R)} E(\rho_2(\mathbf{X}, X_i)), \quad (5)$$

where ρ_2 – suitable metric in the set of fuzzy sets (numbers).

If metric (1) is used, we get

$$\tilde{E}(\mathbf{X}) = \arg \min_{X_i \in F_c(R)} E\left(\left(D_W^\varphi(\mathbf{X}, X_i)\right)^2\right). \quad (6)$$

A fuzzy mean value defined this way is *Aumann-type mean*. This mean optimises convex combination of D_W^φ -distances to components of fuzzy random variable \mathbf{X} .

At any definition of a fuzzy random variable and any definition of metric in the set of fuzzy random sets, the mean of a fuzzy random variable is calculated by (4).

As regards variance of a fuzzy random variable, which is defined in Kwakernaak's sense, a fuzzy value of its variation is calculated as follows:

$$\text{Var}(X)_\alpha = [\inf \text{Var}(X)_\alpha, \sup \text{Var}(X)_\alpha], \quad (7)$$

where values $\inf \text{Var}(X)_\alpha$ are calculated as follows:

$$\inf \text{Var}(X)_\alpha = \frac{\sum_{i=1}^n [\inf E(X)_\alpha - \inf X_{i\alpha}]^2}{n}, \quad (8)$$

but values $\sup \text{Var}(X)_\alpha$ are calculated as follows:

$$\sup \text{Var}(X)_\alpha = \frac{\sum_{i=1}^n [\sup E(X)_\alpha - \sup X_{i\alpha}]^2}{n}, \quad (9)$$

where n denotes the number of assigned α -levels.

If a fuzzy random variable is defined in Puri-Ralescu's sense, the deterministic value of its variance can be calculated using one of these expressions [2, 3, 4]:

$$\text{Var}(X)_\alpha = \sqrt{\sum_{i=1}^n \left[\left(\text{mid}E(X)_\alpha - \text{mid}X_{i\alpha} \right)^2 + W \left(\text{spr}E(X)_\alpha - \text{spr}X_{i\alpha} \right)^2 \right]}, \quad (10)$$

where $\text{mid}E(X)_\alpha$ – the centre of radius of fuzzy mean value $\tilde{E}(\mathbf{X})$ at level α ; $\text{mid}X_{i\alpha}$, $i = 1, \dots, n$, – the centre of radius of fuzzy category \mathbf{X}_i at level α ; $\text{spr}E(X)_\alpha$ – radius of a fuzzy mean value $\tilde{E}(\mathbf{X})$ at level α ; $\text{spr}X_{i\alpha}$, $i = 1, \dots, n$, – radius of a fuzzy category \mathbf{X}_i at level α ; $W \in (0,1]$ – a parameter that evaluates the extent of difference $(\text{spr}U_\alpha - \text{spr}V_\alpha)$ contribution to the general evaluation of the distance.

$$\text{Var}(X)_\alpha = \sqrt{\sum_{i=1}^n \left(\frac{1}{2} [\inf E(X)_\alpha - \inf X_{i\alpha}]^2 + \frac{1}{2} [\sup E(X)_\alpha - \sup X_{i\alpha}]^2 \right)}, \quad (11)$$

where $\inf E(X)_\alpha$ – the lower bound of the interval of possible values of a fuzzy mean value $\tilde{E}(\mathbf{X})$ at level α ; $\inf X_{i\alpha}$, $i = 1, \dots, n$, – the lower bound of the interval of possible values of a fuzzy category \mathbf{X}_i at level α ; $\sup E(X)_\alpha$ – the upper bound of the interval of possible values of a fuzzy mean value $\tilde{E}(\mathbf{X})$ at level α ; $\sup X_{i\alpha}$, $i = 1, \dots, n$, – the upper bound of the interval of possible values of a fuzzy category \mathbf{X}_i at level α .

Common deterministic value of variance $\text{Var}(\mathbf{X})$ can be calculated by averaging values of $\text{Var}(X)_\alpha$ at all α -levels.

More details regarding different issues related to statistical parameters of fuzzy random variables can be found in [11]–[14].

IV. ILLUSTRATIVE EXAMPLES OF CALCULATION OF STATISTICAL PARAMETERS OF A FUZZY RANDOM VARIABLE

Example 1. Let there be set a fuzzy random variable \mathbf{X} with the following fuzzy categories: $\mathbf{X}_1 = (1, 1, 3, 5)$, $\mathbf{X}_2 = (4, 5, 7, 10)$, $\mathbf{X}_3 = (9, 14, 17, 19)$. A random experiment has been conducted, in which a fuzzy category \mathbf{X}_1 has appeared four times, fuzzy category \mathbf{X}_2 has appeared eight times and random category \mathbf{X}_3 has appeared eight times. It is necessary to determine the fuzzy mean value $\tilde{E}(\mathbf{X})$ and fuzzy value of variance $\tilde{Var}(\mathbf{X})$.

Let us calculate appearance frequencies of fuzzy categories in the experiment.

$$f(\mathbf{X}_1) = f_1 = \frac{4}{20} = 0.2, \quad f(\mathbf{X}_2) = f_2 = \frac{8}{20} = 0.4,$$

$$f(\mathbf{X}_3) = f_3 = \frac{8}{20} = 0.4.$$

Let us determine fuzzy numbers:

$$\mathbf{X}_1^* = f_1 * \mathbf{X}_1 = 0.2 * (1, 1, 3, 5) = (0.20, 0.20, 0.60, 1.00)$$

$$\mathbf{X}_2^* = f_2 * \mathbf{X}_2 = 0.4 * (4, 5, 7, 10) = (1.60, 2.00, 2.80, 4.00)$$

$$\mathbf{X}_3^* = f_3 * \mathbf{X}_3 = 0.4 * (9, 14, 17, 19) = (3.60, 5.60, 6.80, 7.60).$$

Graphs of membership functions for fuzzy numbers \mathbf{X}_1^* , \mathbf{X}_2^* and \mathbf{X}_3^* are shown in Fig 3. Using (4), let us calculate boundary values $\inf E(X)_\alpha$ and $\sup E(X)_\alpha$ of a fuzzy mean value $\tilde{E}(\mathbf{X})$.

Initial data and calculation results are given in Table I.

TABLE I
BOUNDARY VALUES OF INTERVALS OF α -CUTS OF FUZZY CATEGORIES \mathbf{X}_1^* , \mathbf{X}_2^* , \mathbf{X}_3^*
AND BOUNDARY VALUES OF INTERVALS OF α -CUTS OF FUZZY MEAN VALUE
 $\tilde{E}(\mathbf{X})$ IN EXAMPLE 1

α	\mathbf{X}_1^*		\mathbf{X}_2^*		\mathbf{X}_3^*		$\tilde{E}(\mathbf{X})$	
	$\inf X_{1\alpha}^*$	$\sup X_{1\alpha}^*$	$\inf X_{2\alpha}^*$	$\sup X_{2\alpha}^*$	$\inf X_{3\alpha}^*$	$\sup X_{3\alpha}^*$	$\inf E(X)_\alpha$	$\sup E(X)_\alpha$
0	0.20	1.00	1.60	4.00	3.60	7.60	1.80	4.20
0.1	0.20	0.96	1.64	3.88	3.80	7.52	1.88	4.12
0.2	0.20	0.92	1.68	3.76	4.00	7.44	1.96	4.04
0.3	0.20	0.88	1.72	3.64	4.20	7.36	2.04	3.96
0.4	0.20	0.84	1.76	3.52	4.40	7.28	2.12	3.88
0.5	0.20	0.80	1.80	3.40	4.60	7.20	2.20	3.80
0.6	0.20	0.76	1.84	3.28	4.80	7.12	2.28	3.72
0.7	0.20	0.72	1.88	3.16	5.00	7.04	2.36	3.64
0.8	0.20	0.68	1.92	3.04	5.20	6.96	2.44	3.56
0.9	0.20	0.64	1.96	2.92	5.40	6.88	2.52	3.48
1	0.20	0.60	2.00	2.80	5.60	6.80	2.60	3.40

TABLE II
CALCULATION RESULTS FOR VALUES $\inf Var(X)$ AND $\sup Var(X)$ IN EXAMPLE 2

α	$\inf E(X)_\alpha$	$\inf X_{1\alpha}^*$	$\inf X_{2\alpha}^*$	$\inf X_{3\alpha}^*$	$\sup E(X)_\alpha$	$\sup X_{1\alpha}^*$	$\sup X_{2\alpha}^*$	$\sup X_{3\alpha}^*$	$\inf Var(X)_\alpha$	$\sup Var(X)_\alpha$
0	1.80	0.20	1.60	3.60	4.20	1.00	4.00	7.60	1.95	7.28
0.1	1.88	0.20	1.64	3.80	4.12	0.96	3.88	7.52	2.19	7.20
0.2	1.96	0.20	1.68	4.00	4.04	0.92	3.76	7.44	2.45	7.12
0.3	2.04	0.20	1.72	4.20	3.96	0.88	3.64	7.36	2.72	7.05
0.4	2.12	0.20	1.76	4.40	3.88	0.84	3.52	7.28	3.01	6.98
0.5	2.20	0.20	1.80	4.60	3.80	0.80	3.40	7.20	3.31	6.91
0.6	2.28	0.20	1.84	4.80	3.72	0.76	3.28	7.12	3.62	6.84
0.7	2.36	0.20	1.88	5.00	3.64	0.72	3.16	7.04	3.96	6.77
0.8	2.44	0.20	1.92	5.20	3.56	0.68	3.04	6.96	4.30	6.71
0.9	2.52	0.20	1.96	5.40	3.48	0.64	2.92	6.88	4.66	6.66
1	2.60	0.20	2.00	5.60	3.40	0.60	2.80	6.80	5.04	6.59

The graph of membership function of fuzzy mean value $\tilde{E}(X)$ can be seen in Fig. 3.

Example 2. Fuzzy value of variance $V\tilde{ar}(X)$ has to be calculated at initial data of Example 1. Let us make calculations using (7, 8, 9). Initial data and calculation results are given in Table II.

The graph of membership function of fuzzy variance $V\tilde{ar}(X)$ is shown in Fig. 3.

Example 3. Using initial data of Example 1, it is necessary to calculate a deterministic value of dispersion $Var(X)$ by (11).

Initial data and calculation results are given in Table III. By averaging last column values in Table III, we obtain $Var(X) = 3.92$.

V. RESULT ANALYSIS

As regards a fuzzy mean value of a fuzzy random variable, it is calculated by (4) not taking into account in what sense that variable has been defined. If a relevant fuzzy random variable is defined in Kwakernaak's sense, the variance of this variable has to be evaluated in the fuzzy form (see Example 2). This directly follows from the fact that such a fuzzy random variable is defined as a fuzzy perception of underlying it real random variable.

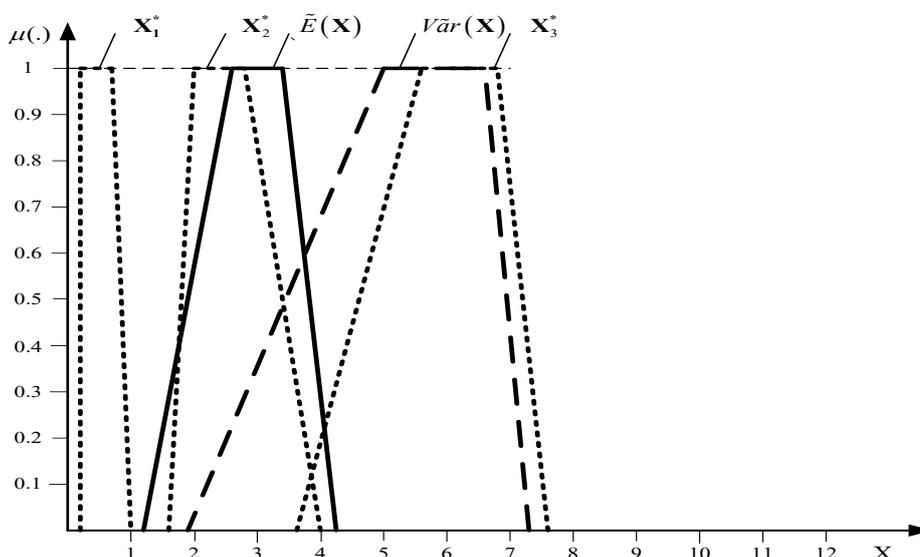


Fig. 3. Graphs of membership functions for fuzzy numbers X_1^* , X_2^* , X_3^* in Example 1, fuzzy mean value $\tilde{E}(X)$ (Example 1) and fuzzy variance $V\tilde{ar}(X)$ (Example 2).

TABLE III
CALCULATION RESULTS FOR VALUES $Var(X)_\alpha$ IN EXAMPLE 3

α	inf $E(X)_\alpha$	inf $X_{1\alpha}^*$	inf $X_{2\alpha}^*$	inf $X_{3\alpha}^*$	sup $E(X)_\alpha$	sup $X_{1\alpha}^*$	sup $X_{2\alpha}^*$	sup $X_{3\alpha}^*$	$Var(X)_\alpha$
0	1.80	0.20	1.60	3.60	4.20	1.00	4.00	7.60	3.72
0.1	1.88	0.20	1.64	3.80	4.12	0.96	3.88	7.52	3.75
0.2	1.96	0.20	1.68	4.00	4.04	0.92	3.76	7.44	3.79
0.3	2.04	0.20	1.72	4.20	3.96	0.88	3.64	7.36	3.83
0.4	2.12	0.20	1.76	4.40	3.88	0.84	3.52	7.28	3.87
0.5	2.20	0.20	1.80	4.60	3.80	0.80	3.40	7.20	3.91
0.6	2.28	0.20	1.84	4.80	3.72	0.76	3.28	7.12	3.96
0.7	2.36	0.20	1.88	5.00	3.64	0.72	3.16	7.04	4.01
0.8	2.44	0.20	1.92	5.20	3.56	0.68	3.04	6.96	4.06
0.9	2.52	0.20	1.96	5.40	3.48	0.64	2.92	6.88	4.11
1	2.60	0.20	2.00	5.60	3.40	0.60	2.80	6.80	4.16

This, in its turn, assumes that the parameters of such a variable are defined as fuzzy perceptions of the parameters of the original random variable.

To analyse the deterministic variance, let us consider an example that follows. Let there be a hypothetical system of pattern recognition that, based on certain features, automatically ascribes it to one of the fuzzy categories of age. Having this kind of system, random experiment is as follows: out of a population of individuals, a subset of individuals is randomly chosen (a sample of individuals is created). Each individual in the sample is assigned a fuzzy category of age by evaluating the age not numerically but just by the features observed. Since this fuzzy random variable is explicitly defined in Puri-Ralescu's sense, its variance has to be calculated as deterministic number.

A detailed analysis of both kinds of variance is provided in [11]–[14]. The deterministic value of variance is only determined by the form of membership function of fuzzy categories and “the distance” between fuzzy categories on the relevant attribute measurement scale. As is figuratively pointed in [12], scalar variance measures the variability of membership functions, but not the variability of the underlying fixed-point number. The main contribution to the deterministic value of variation is made by “the distance” between relevant fuzzy categories. The spread of values within each fuzzy category and within fuzzy mean value only slightly contributes to the overall variance evaluation.

As a positive feature of deterministic values of variance, just their deterministic nature can be considered. If there are two fuzzy random variables, and the values of their variance are represented in the deterministic form, one can easily compare, if necessary also numerically, the extent of spread of fuzzy components of both variables. It is evident that the result of such a comparison will be correct under the condition that the same metric was used for calculating both values.

One shortcoming of deterministic evaluation of variation is that the evaluated variance value depends on the metric on whose basis that value was calculated.

The calculation of fuzzy values of variance requires a considerable volume of calculations, especially in the case of a large number of components of relevant fuzzy random variable.

The advantage of the approach is a common procedure for calculating the fuzzy value of variation that is based on a simple metric in the set of real numbers. However, if for two fuzzy random variables fuzzy variance values are calculated, comparison and analysis of the results can cause difficulties. At best, we can only make a conclusion that the extent of spread of components of one fuzzy random variable is greater (or smaller) than the spread of components of other fuzzy random variable, without numerical result of such a comparison.

Another specific feature of fuzzy evaluation of variance is that we cannot obtain interval values at all α -levels that are interesting to us. This can be an advantage, especially in the case, when membership functions of the initial categories and membership function of the fuzzy mean value have a complicated form.

In common case, it is impossible to give preference to either of the considered techniques of calculating variance values for fuzzy random variables. As stated in [11], those techniques either serve to different purposes or represent different models of the phenomenon under consideration, as well as different observation environments.

VI. CONCLUSION

Over the past decades, an intensive development of methods offering successful processing of uncertain initial data, namely, interval and fuzzy data, has been seen. The most common type of initial data uncertainty is observed when these methods are represented in the fuzzy form. Here, two types of uncertainty, randomness and fuzziness operate at the same time. Formal expression of this kind of combination is associated with the concept of a *fuzzy random variable*. The situation becomes more complicated due to the fact that in real-world practice two conceptual interpretations of such variables are widely used, i.e., Kwakernaak's interpretation and Puri-Ralescu's interpretation. Those different interpretations cause the need to develop specific calculation algorithms for determining statistical parameters of initial fuzzy random variables.

Other points should also be mentioned that make the initial acquaintance with fuzzy random variables difficult: the lack of united system of basic concepts and denotations, a large number of complicated theoretical works and essential shortage of works oriented towards an ordinary user.

From the theoretical point of view, no preference can be given to either alternative definition of a fuzzy random variable. Both definitions are correct; the difference between them is only related to conceptual bases of each definition. From the practical point of view, the choice between the alternatives is determined by the essence of the task to be solved and required/desirable representation of the results. If the initial data are apriori represented in the form of fuzzy categories, the Puri-Ralescu's approach should be used. Whereas, if it is necessary to analyse the character of variance changes by α -levels, Kwakernaak's approach is preferable. Anyway, the choice of either approach has to be based on the essence of the task being solved, as well as its purpose and type of initial data.

Further theoretical and practical works in this field should be directed towards the development of methods of correct dissemination of the results obtained in a sample, to the corresponding population. This is the most topical task of statistics with fuzzy initial data.

REFERENCES

- [1] L. Zadeh, “The concept of a linguistic variable and its application to approximate Reasoning”, *Inform. Sci.*, vol. 8, no. 3, pp. 199–249, 1975. [https://doi.org/10.1016/0020-0255\(75\)90036-5](https://doi.org/10.1016/0020-0255(75)90036-5)
- [2] H. Kwakernaak, “Fuzzy Random Variables—I. Definitions and Theorems”, *Information Sciences*, vol. 15, no. 1, pp. 1–29, 1978. [https://doi.org/10.1016/0020-0255\(78\)90019-1](https://doi.org/10.1016/0020-0255(78)90019-1)
- [3] H. Kwakernaak, “Fuzzy Random Variables—II. Algorithms and Examples for the Discrete Case”, *Information Sciences*, vol. 17, no. 3, pp. 253–278, 1979. [https://doi.org/10.1016/0020-0255\(79\)90020-3](https://doi.org/10.1016/0020-0255(79)90020-3)
- [4] R. Kruse, “On the Variance of Random Sets”, *Journal of Mathematical Analysis and Applications*, vol. 122, no. 2, pp. 469–473, 1987. [https://doi.org/10.1016/0022-247X\(87\)90277-0](https://doi.org/10.1016/0022-247X(87)90277-0)

- [5] R. Kruse, and K. D. Meyer, *Statistics with Vague Data*, D. Reidel Publishing Company, Dordrecht-Boston-Lancaster-Tokyo, 1987. <https://doi.org/10.1007/978-94-009-3943-1>
- [6] M. L. Puri, and D. A. Ralescu, "Fuzzy random variables", *Journal of Mathematical Analysis and Applications*, vol. 114, no. 2, pp. 409–422, 1986. [https://doi.org/10.1016/0022-247X\(86\)90093-4](https://doi.org/10.1016/0022-247X(86)90093-4)
- [7] M. A. Gil, "Fuzzy Random Variables à la Kruse & Meyer and à la Puri & Ralescu: Key Differences and Coincidences", *Studies in Computational Intelligence*, pp. 21–29, 2017. <https://doi.org/10.1007/978-3-319-67789-7>
- [8] C. Bertoluzza, N. Corall, and A. Salas, "On a new class of distance between fuzzy numbers", *Mathware & Soft Computing*, vol. 2, pp. 71–84, 1995.
- [9] A. Blanco-Fernández, M. R. Casals, A. Colubi, N. Coral, M. Garsía-Bárcana, M. A. Gil, G. González-Rodríguez, M. T. Lopes, M. A. Lubiano, M. Montenegro, A. B. Ramos-Guajardo, S. de la Rosa de Sáa, B. Sinova, "A distance-based statistical analysis of fuzzy number-valued data", *Int. J. Approx. Reas.*, vol. 55, no. 7, pp. 1487–1501, 2014. <https://doi.org/10.1016/j.ijar.2013.09.020>
- [10] A. Colubi, R. Coppi, P. D'Urso, M. A. Gil, "Statistics with fuzzy random variables," *METRON- Int. Journal of Statistics*, vol. 3, pp. 277–303, 2007.
- [11] I. Couso, D. Dubois, S. Montes, and L. Sánchez, "On various definitions on the variance of a fuzzy random variable", *5th Int. Symposium on Imprecise Probability: Theories and Applications*, Prague, Czech Republic, 10 p. 2007.
- [12] I. Couso, and D. Dubois, "On the variability of the concept of variance for fuzzy random variables," *IEEE Trans. Fuzzy Systems*, vol. 17, no. 5, pp. 1070–1080, 2009. <https://doi.org/10.1109/TFUZZ.2009.2021617>
- [13] I. Couso, and D. Dubois, "Statistical reasoning with set-valued information: Ontic vs. epistemic views," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1502–1518, 2014. <https://doi.org/10.1016/j.ijar.2013.07.002>
- [14] M. A. Gil, "Statistics with fuzzy data by using random fuzzy sets", *Proceedings 58th World Statistical Congress*, Dublin, 2011, pp. 1283–1290.

Oļegs Užga-Rebrovs is a Leading Researcher of the Information and Communication Technologies Research Centre at Rezekne Academy of Technologies (Latvia). He received his Doctoral degree in Information Systems from Riga Technical University in 1994. His research interests include different approaches to processing incomplete, uncertain and fuzzy information, in particular, fuzzy sets theory, rough set theory as well as fuzzy classification and fuzzy clustering techniques and their applications in bioinformatics. E-mail: ushga@ru.lv

Galina Kulešova is a Researcher of the Faculty of Computer Science and Information Technology at Riga Technical University (Latvia). She received her *M. Sc.* degree in Decision Support Systems from Riga Technical University. Current research interests include artificial neural networks, data mining, ontology engineering, classification methods and bioinformatics. E-mail: galina.kulesova@rtu.lv