

Fuzzy Robust Estimates of Location and Scale Parameters of a Fuzzy Random Variable

Oleg Uzhga-Rebrov

Rezekne Academy of Technologies

Rezekne, Latvia

rebrows@tvnet.lv

Galina Kuleshova

Dept. of Modelling and Simulation

Riga Technical University

Riga, Latvia

galina.kulesova@rtu.lv

Abstract - A random variable is a variable whose components are random values. To characterise a random variable, the arithmetic mean is widely used as an estimate of the location parameter, and variation as an estimate of the scale parameter. The disadvantage of the arithmetic mean is that it is sensitive to extreme values, *outliers* in the data. Due to that, to characterise random variables, robust estimates of the location and scale parameters are widely used: the median and median absolute deviation from the median. In real situations, the components of a random variable cannot always be estimated in a deterministic way. One way to model the initial data uncertainty is to use fuzzy estimates of the components of a random variable. Such variables are called *fuzzy random variables*. In this paper, we examine fuzzy robust estimates of location and scale parameters of a fuzzy random variable: fuzzy median and fuzzy median of the deviations of fuzzy component values from the fuzzy median.

Keywords - *fuzzy median, fuzzy median of absolute deviations from the fuzzy median, fuzzy random variable, random variable.*

I. INTRODUCTION

Let there be a population or sample containing n objects. These objects are characterised by multiple values of the estimating attribute $\mathbf{X} = \{x_1, \dots, x_n\}$. Then the set of values \mathbf{X} can be correctly considered as a random variable, whose elements are values $x_i \in \mathbf{X}$.

In standard statistics, a random variable \mathbf{X} can be characterised (or described) by two parameters:

- location parameter $E(\mathbf{X})$, which is the arithmetic mean of components

$$E(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

- scale parameter (variation) $V(\mathbf{X})$ that describes deviations of components $x_i \in \mathbf{X}$ from the mean value, $E(\mathbf{X})$

$$V(\mathbf{X}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - E(\mathbf{X}))^2 \quad (2)$$

The disadvantage of the location parameter (1) is that its value is greatly influenced by extreme data values, *outliers*. To mitigate such unwanted influences, various robust estimates of location and scale parameters have been proposed that are insensitive to the presence of outliers in the data.

The simplest robust estimator of the location parameter of a random variable is its median. The median value is determined as follows. Let us have a random variable $\mathbf{X} = \{x_1, \dots, x_n\}$.

Let us order the values $x_i \in \mathbf{X}$ in non-decreasing order: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n-1)} \leq x_{(n)}$ where the subscript in parentheses represents the component number in their ordered sequence. Then the median of the relevant random variable is defined as

$$\text{med}(\mathbf{X}) = x_{(n+1)/2}, \quad n - \text{even}; \quad (3a)$$

Online ISSN 2256-070X

<https://doi.org/10.17770/etr2021vol2.6566>

© 2021 Oleg Uzhga Rebrov, Galina Kuleshova. Published by Rezekne Academy of Technologies. This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

$$\text{med}(\mathbf{X}) = \frac{1}{2}(x_{(k)} + x_{(k+1)}), \quad n = 2k - \text{odd}. \quad (3b)$$

In this case, the median of absolute deviations from the median is used as a robust estimator of the scale parameter

$$\text{mad}(\mathbf{X}) = \text{med} \left| x_{(i)} - \text{med}(\mathbf{X}) \right|, \quad i = 1, \dots, n. \quad (4)$$

Let us consider a simple illustrative example. Let there be $\mathbf{X} = \{3, 7, 14, 32\}$. Using expression (3b), we can calculate the median value for these data

$$\text{med}(\mathbf{X}) = \frac{1}{2}(x_{(2)} + x_{(3)}) = \frac{1}{2}(7 + 14) = 10.50.$$

For comparison, using expression (1), let us calculate the arithmetic mean value for a given random variable

$$E(\mathbf{X}) = \frac{1}{4}(3 + 7 + 14 + 32) = 14.00.$$

Obviously, the mean is affected by the extreme value $x_{(4)} = 32$, which could potentially be an outlier in these data.

Let us calculate the values of the absolute deviations of the initial data from the median. We have the following expression:

$\{x_{(i)} - \text{med}(\mathbf{X})\} = \{7.50, 3.50, 3.50, 21.50\}$. Let us arrange these values in non-decreasing order: 3.50, 3.50, 7.50, 21.50. By expression (4), we have

$$\text{mad}(\mathbf{X}) = \text{med}\{3.50, 3.50, 7.50, 21.50\} = \frac{1}{2}(3.50 + 7.50) = 5.50$$

In many real-world situations, it is impossible to specify precise numerical values for the relevant attribute for a population or sample. Therefore, there is a natural need to set the attribute values in some suitable uncertain form. One of the widely used options for modelling uncertainties in the initial data is to use fuzzy numbers as estimates of attribute values.

In the general case, a fuzzy number represents a fuzzy set on the scale of real numbers. To set a fuzzy number, it is necessary to determine its support (interval) on a certain measurement scale x . For each point x inside the support, the value of the function of membership $\mu(x)$ of the value x at this point to a fuzzy number $\tilde{\mathbf{X}}$ must be known. The value x for which $\mu(x) = 1$, is called the core of the fuzzy number $\tilde{\mathbf{X}}$. If the maximum value of the membership function of a fuzzy number is 1, such a number is called a *normal* fuzzy number.

The membership function of a fuzzy number $\tilde{\mathbf{X}}$ can be set analytically, graphically, or in the form of calculated values for intervals at any of its α -levels. In this paper, we will use normal fuzzy numbers in the simplest form,

namely, triangular fuzzy numbers. Fig. 1 shows a graph of the membership function of a conditional normal triangular fuzzy number.

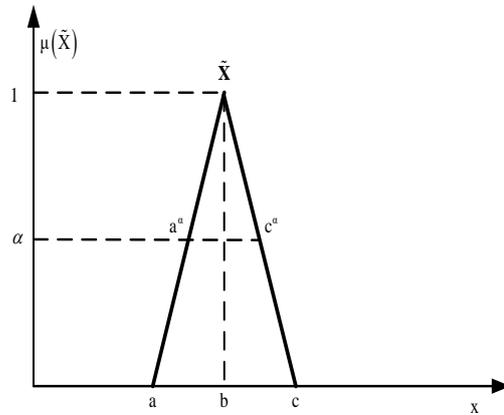


Fig.1. Graphic representation of a normal triangular fuzzy number $\tilde{\mathbf{X}}$.

The support of a fuzzy number $\tilde{\mathbf{X}}$ is an interval $[a, c]$, point b corresponds to its core, for which $\mu(\tilde{\mathbf{X}}) = 1$. If we set some value α of the membership function, to this number corresponds an interval (the interval $[a^\alpha, c^\alpha]$ in Fig. 1). This interval is called the α -cut of the fuzzy number $\tilde{\mathbf{X}}$. Evidently, when $\alpha = 0$, the α -cut is the support of the fuzzy number $\tilde{\mathbf{X}}$, and the value $\alpha = 1$ corresponds to the core of this fuzzy number.

Using the notation shown in Fig. 1, a fuzzy number $\tilde{\mathbf{X}}$ can be given in the form of a triplet of numbers $\tilde{\mathbf{X}} = (a, b, c)$, where a and c are the boundaries of the support $\tilde{\mathbf{X}}$, b is the value x that corresponds to the core $\tilde{\mathbf{X}}$. We will use this notation in the subsequent sections of the paper.

A random variable whose components are fuzzy numbers is called a *fuzzy random variable*. There are two alternative formal definitions of the concept of a fuzzy random variable. The first definition is presented in [4, 5]. The second definition is presented in [6]. The difference between these definitions is only in the interpretation of the components of a fuzzy random variable. A detailed analysis of these differences is provided in [7]. Various aspects of the theory and practice of fuzzy random variables can be found in [8- 12]. According to the authors of formal definitions, the first type of such variables is called fuzzy random variables in Kwakernaak's sense, the second type is called fuzzy random variables in Puri/Ralescu's sense.

In the context of fuzzy statistics, alternative definitions of a fuzzy random variable lead to different definitions of the variation of such a variable.

For the purposes of representing and analysing fuzzy robust parameters of the location and scale of a fuzzy

random variable, it does not matter in this article how this variable is formally defined.

II. CALCULATION OF FUZZY PARAMETERS OF A FUZZY RANDOM VARIABLE

For the purposes of comparative analysis, we first present an expression that allows us to calculate the interval of the fuzzy mean value of a fuzzy random variable at the level α

$$E(X)_\alpha = [E(\inf x_\alpha), E(\sup x_\alpha)] \quad (5)$$

It follows from this expression that the lower limit of the interval of the fuzzy number $\tilde{E}(X)$ at the level α is calculated as the average value $\inf X_\alpha$, i.e., as the average value of the corresponding values on the ascending parts of the membership functions of fuzzy numbers $\tilde{X}_i, i = 1, \dots, n$. In its turn, the upper limit of the interval a fuzzy number $\tilde{E}(X)$ at the level α is calculated as the average value $\sup X_\alpha$, that is, as the average value of the corresponding values on the falling parts of the membership functions of fuzzy numbers $\tilde{X}_i, i = 1, \dots, n$.

The fuzzy mean $\tilde{E}(\tilde{X})$ has the same disadvantage as its crisp counterpart $E(X)$, namely, it is sensitive to outliers in the data. The fuzzy median is devoid of this synonym drawback and can be used as a fuzzy robust estimator of the location parameter of a fuzzy random variable. The theoretical foundations of fuzzy medians are discussed in [12, 13].

Let us introduce the following definitions.

Definition 1 [13]. Mapping $F_c(R) \times F_c(R) \rightarrow [0, \infty)$, such that, for $\tilde{U}, \tilde{V} \in F_c(R)$

$$\rho_1(\tilde{U}, \tilde{V}) = \left\| s_{\tilde{U}} - s_{\tilde{V}} \right\| = \frac{1}{2} [(0,1) | \inf \tilde{U}_\alpha - \inf \tilde{V}_\alpha | + (6) \\ + | \sup \tilde{U}_\alpha - \sup \tilde{V}_\alpha |] d\alpha$$

is called *the distance of norm-1 between fuzzy numbers*. In this definition, $F_c(R)$ is a set of fuzzy sets (numbers) defined on a real scale R .

Definition 2 [13]. For a given probability space (Ω, A, P) and associated fuzzy random variable \tilde{X} , the *median* (or medians) of the distribution \tilde{X} is a fuzzy number (or fuzzy numbers).

$$E(\rho_1, M\tilde{e}(\tilde{X})) = \min_{\tilde{U} \in F_c(R)} E(\rho_1(\tilde{X}, \tilde{U})), \quad (7)$$

whenever such an expectation exists. In Definition (2), $F_c(R)$ is a set of all fuzzy sets defined on a real scale R , ρ_1 is a metric of norm-1 – from Definition (1).

The following definition is presented in [15] in the form of a theorem with a strong proof.

Definition 3. For a given probability space (Ω, A, P) and a fuzzy random variable \tilde{X} associated with it, for any value $\alpha \in [0, 1]$, we have a fuzzy number $M\tilde{e}(\tilde{X}) \in F_c(R)$ such that

$$M\tilde{e}(\tilde{X})_\alpha = [Me(\inf X_\alpha), Me(\sup X_\alpha)] \quad (8)$$

where, if $Me(\inf X_\alpha)$ and $Me(\sup X_\alpha)$ are not unique and following customary agreements, $Me(\inf X_\alpha)$ is selected as the midpoint of the median interval from $\inf X_\alpha$; $Me(\inf X_\alpha)$ is selected as the midpoint of the median interval from $\inf X_\alpha$; $Me(\sup X_\alpha)$ is selected as the midpoint of the median interval from $\sup X_\alpha$.

Note that the scheme for calculating the fuzzy median $M\tilde{e}(\tilde{X})$ is similar to that for calculating the fuzzy mean $\tilde{E}(\tilde{X})$. The difference is that values $\inf E(X)_\alpha$ and $\sup E(X)_\alpha$ are calculated as averages of values $\inf X_{i\alpha}$, $\sup X_{i\alpha} i = 1, \dots, n$, but values $\inf(Me(X_\alpha))$ and $\sup(Me(X_\alpha))$, are calculated as medians of values $\inf X_\alpha, \sup X_\alpha$.

In the previous section, expressions were presented for calculating the median value on the set of real-valued initial data (expressions (3, a, b)). Let us modify these expressions so that they can be used to calculate the fuzzy median $M\tilde{e}(\tilde{X})$.

Let a fuzzy random variable $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_i, \dots, \tilde{X}_n)$ be given. First, sets of values $\{\inf X_{i\alpha} / i = 1, \dots, n\}$, $\{\sup X_{i\alpha} / i = 1, \dots, n\}$. Then the values $\inf X_{i\alpha}, \sup X_{i\alpha}$ must be ordered in order of increasing their values. The value $Me(\inf X_\alpha)$ is defined as

$$Me(\inf X_\alpha) = \inf X_{\left(\frac{n+1}{2}\right)}, \quad (9a)$$

where n is an odd number.

$$Me(\inf X_\alpha) = \frac{1}{2} \left(\inf X_{\left(\frac{n}{2}\right)} + \inf X_{\left(\frac{n+1}{2}\right)} \right), \quad (9b)$$

where n is an even number.

Similarly,

$$Me(\sup X_\alpha) = \sup X_{\left(\frac{n+1}{2}\right)}, \quad (10a)$$

where n is an odd number.

$$Me(\sup X_\alpha) = \frac{1}{2} \left(\sup X_{\left(\frac{n}{2}\right)} + \sup X_{\left(\frac{n+1}{2}\right)} \right) \quad (10b)$$

where n - is an even number.

Expressions (9a, 9b) and (10a, 10b) are formal representations of the conventions in Definition (3).

Let us consider a simple illustrative example. A fuzzy random variable $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3, \tilde{X}_4)$ is given, where, $\tilde{X}_1 = (1, 3, 5)$, $\tilde{X}_2 = (4, 7, 10)$, $\tilde{X}_3 = (9, 14, 19)$, $\tilde{X}_4 = (30, 32, 34)$. It is necessary to determine the fuzzy median of this fuzzy random variable.

Let us derive calculation expressions for determining the intervals of fuzzy components $\tilde{X}_1, \tilde{X}_2, \tilde{X}_3,$ and \tilde{X}_4 at levels α .

- fuzzy component \tilde{X}_1 :

$$\inf X_{1\alpha} = 1 + (3 - 1)\alpha = 1 + 2\alpha$$

$$\sup X_{1\alpha} = 5 - (5 - 3)\alpha = 5 - 2\alpha ;$$

- fuzzy component \tilde{X}_2 :

$$\inf X_{2\alpha} = 4 + (7 - 4)\alpha = 4 + 3\alpha$$

$$\sup X_{2\alpha} = 10 - (10 - 7)\alpha = 10 - 3\alpha ;$$

- fuzzy component \tilde{X}_3 :

$$\inf X_{3\alpha} = 9 + (19 - 14)\alpha = 9 + 5\alpha$$

$$\sup X_{3\alpha} = 19 - (19 - 14)\alpha = 19 - 5\alpha ;$$

- fuzzy component \tilde{X}_4 :

$$\inf X_{4\alpha} = 30 + (32 - 30)\alpha = 30 + 2\alpha$$

$$\sup X_{4\alpha} = 34 - (34 - 32)\alpha = 34 - 2\alpha .$$

Given the nature of the relevant fuzzy numbers, there is no need to perform calculations at intermediate α levels. It is enough to perform calculations at the levels $\alpha = 0$ and $\alpha = 1$.

$\alpha = 0$: $\{\inf X_{i0}\} = \{1, 4, 9, 30\}$. By expression (9b),

$$Me(\inf X_0) = \frac{1}{2}(4 + 9) = 6.50 \quad \{\sup X_{i0}\} = \{5, 10, 19, 34\} .$$

By expression (10b), $Me(\sup X_0) = \frac{1}{2}(10 + 19) = 14.50$.

$$\alpha = 1 : Me(X_1) = \frac{1}{2}(7 + 14) = 10.5 .$$

Thus, the fuzzy median in our example is $Me(\tilde{X}) = (6.50, 10.50, 14.50)$. The graph of the membership function together with the graphs of the membership functions of the original fuzzy numbers $\tilde{X}_i, i = 1, 2, 3, 4$ is shown in Fig. 2. For comparison, this figure shows a graph of the fuzzy average value $E(\tilde{X})$ calculated by expression (5).

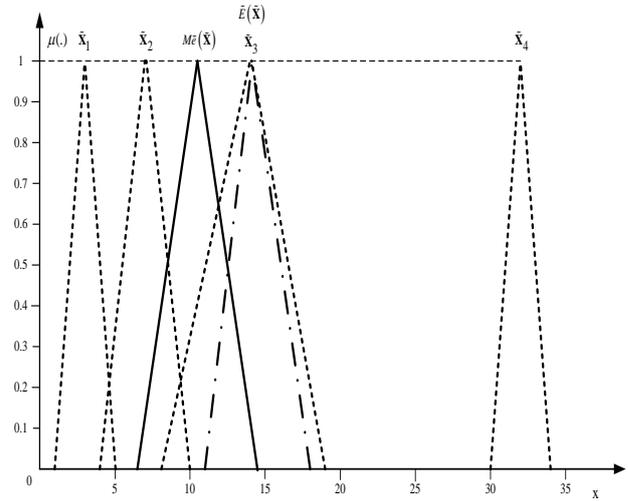


Fig.2. Graphs of membership functions of fuzzy components $\tilde{X}_i, i = 1, 2, 3, 4$, fuzzy median $Me(\tilde{X})$ and fuzzy mean $E(\tilde{X})$.

By analogy with the median of absolute deviations from the median of a deterministic random variable, we define the fuzzy median of fuzzy deviations from the fuzzy median.

$$ma(X)_\alpha = \left[\inf \left\{ \left| X_{(i)\alpha} - me(X)_\alpha \right| \right\}, \sup \left\{ \left| X_{(i)\alpha} - me(X)_\alpha \right| \right\} \right], \quad i = 1, \dots, n \quad (11)$$

Let us define a fuzzy value $ma(\tilde{X})$ for the fuzzy random variable \tilde{X} presented above. The calculations will be performed at the levels $\alpha = 0$ and $\alpha = 1$.

$\alpha = 0$:

$$\left| X_{10} - me(X)_0 \right| = \left| [1.00, 5.00] - [6.50, 14.50] \right| = [5.50, 9.50] ;$$

$$\left| X_{20} - me(X)_0 \right| = \left| [4.00, 10.00] - [6.50, 14.50] \right| = [2.50, 4.50] ;$$

$$\left| X_{30} - me(X)_0 \right| = \left| [9.00, 19.00] - [6.50, 14.50] \right| = [2.50, 4.50] ;$$

$$\begin{aligned} |X_{40} - \text{me}(X)_0| &= |[30.00, 34.00] - [6.50, 14, 50]| = \\ &= [19.50, 23.50] \end{aligned}$$

Let us arrange the lower bounds of the resulting intervals in non-decreasing order: 2.50, 2.50, 5.50, 19.50 . The median of these values is:
 $\text{med} = \frac{1}{2}(2.50 + 5.50) = 4.00$.

Let us arrange the upper bounds of the resulting intervals in non-decreasing order: 4.50, 4.50, 9.50, 23.50 . The median of these values is
 $\text{med} = \frac{1}{2}(4.50 + 9.50) = 7.00$.

It follows that the support of the desired fuzzy number $M\tilde{a}(\tilde{X})$ is $\text{ma}(X)_0 = [4.00, 7.00]$

$$\alpha = 1 :$$

At this level, we are dealing with deterministic numbers. The absolute values of the differences between the point values of the components and the point value of the median form the following sequence: 7.50, 3.50, 3.50, 21.50 . Let us sort these numbers in a non-decreasing sequence: 3.50, 3.50, 7.50, 21.50 . The median of this sequence is $\text{med} = \frac{1}{2}(3.50 + 7.50) = 5.50$. Thus, we have the following fuzzy number representing the fuzzy median of the absolute deviations of fuzzy components from the fuzzy median $M\tilde{a}(\tilde{X}) = (4.00, 5.50, 7.00)$. The graph of the membership function of this fuzzy number, together with the initial and previous resulting data, is shown in Fig. 3.

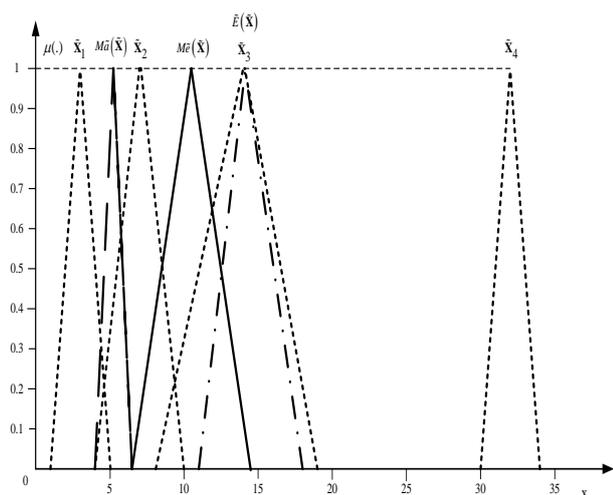


Fig.3. Graphs of membership functions of fuzzy components \tilde{X}_i , $i = 1, 2, 3, 4$, fuzzy median $M\tilde{a}(\tilde{X})$, fuzzy median of deviations from fuzzy median $M\tilde{a}(\tilde{X})$ and fuzzy mean $\tilde{E}(\tilde{X})$.

III. RESULTS AND CONCLUSIONS

In this paper we consider extensions of the concept of median and median of absolute deviations from the median as robust estimators of parameters of a random variable to fuzzy initial data. In standard statistics, the mean and variation of the values of a random variable are widely used as parameters for location and scale. However, the mean has a significant drawback as an estimator of the location parameter :it is sensitive to extreme values in the data. Extreme values in the data can cause the estimated mean to shift towards large or small data values. In other words, in such situations, the average value cannot serve as an adequate estimator of the location parameter.

To reduce or completely avoid the impact of outliers in the data on the estimation of random variable parameters, an estimator is desirable whose values are not affected by large or small extreme values in the data. In other words, a robust estimator of location parameter is needed. Nowadays, different variants of such estimators are available. One of the most common options is to use the median of a random variable as an estimator of the placement parameter of this random variable.

The use of the median makes it necessary to use the corresponding estimator of the scale parameter. The median of absolute deviations from the median is used as such an estimator.

This paper presents an extension of the concepts of the median and median of absolute deviations from the median to the case of fuzzy random variables, that is, random variables whose components are given in a fuzzy form. The theoretical part of the work is supplemented with an illustrative example. Based on this example, an unambiguous conclusion can be made that the introduced fuzzy estimator of location parameters has a high degree of robustness. The corresponding fuzzy scale parameter has the advantage of simplicity of calculation. Calculations of the fuzzy median of absolute deviations from the fuzzy median are significantly simplified for normal triangular fuzzy numbers, which was demonstrated by the example in this work. With more complex forms of membership functions, calculations must be performed at a larger number of α levels. This only increases the volume of the calculations performed, but not their complexity. Calculating the fuzzy value of the variation is much more complicated and requires a lot of computational costs.

Summarizing, we can make a reasonable conclusion that the use of the fuzzy analogues of robust estimators of location and scale parameters presented in this work can be confidently recommended for processing fuzzy initial data.

REFERENCES

- [1] F. R. Hampel, E.M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, Linear Networks and Systems. New York, John Wiley&Sons, 1986.
- [2] P. J. Huber. Robust Statistics. New York, John Wiley&Sons, 1981.

- [3] O.Uzhga-Rebrov, Managing Uncertainties. Part 2. Modern Methods of Probabilistic Reasoning, Rezekne, RA Izdevniecība, 2007. Ужга-Ребров О. И. Управление неопределённостями. Часть 2. Современные методы вероятностного вывода. Rēzekne, RA izdevniecība, 2007.
- [4] H. Kwakernaak, "Fuzzy Random Variables – I. Definitions and Theorems", Information Sciences, 15, 1978, pp. 1 – 29.
- [5] H. Kwakernaak, "Fuzzy Random Variables –II. Definitions and Examples for the Discrete Case", Information Sciences, 17, pp. 253-278, 1979.
- [6] M. L. Puri. and D. A. Ralescu, "Fuzzy random variables", J. Math. Anal. Appl., 114, pp. 409 – 422, 1986.
- [7] A. F. Shapiro, Implementing Fuzzy Random Variables – Some Preliminary Observations. ARC Proceedings, August 1-4, 2012.
- [8] A. Colubi, R. Coppi, P. D'Urso and M. A. Gil, "Statistics with fuzzy random variables". METRON – International Journal of Statistics, Vol. LXV, No. 3, pp. 277 – 303, 2007.
- [9] R. Coppi, M. A. Gil and H.A.L. Kiers, "The fuzzy approach to statistical analysis", Computational Statistics & Data Analysis. Vol. 51, Issue 1, pp. 1 – 14, 2006.
- [10] P. D'Urso and M. A. Gil, "Fuzzy Statistical Analysis: Methods and Applications", METRON, 71, pp. 197-199, 2013.
- [11] M.A. Gil, M. López-Díaz and D.A. Ralescu, "Overview of the Development of Fuzzy Random Variables". Fuzzy Sets and Systems, 157, (19), pp. 2546 – 2557, 2005.
- [12] B. Sinova, M.R. Casals and M.A. Gil, "Central tendency for symmetric random fuzzy numbers", Information Sciences, 278, pp. 599 – 613, 2014.
- [13] B. Sinova, M. A. Gil, A. Colubi, and S. van Aelst. "The median of random fuzzy number. The 1-norm distance approach". Fuzzy Sets and Systems, 200, 2012, pp. 99 – 115.
- [14] I. Couso, S. Montes and P. Gil, "The necessity of the strong α - cuts of a fuzzy sets", Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 9, No. 2, pp. 249 – 262. 2001.