TWMS J. Pure Appl. Math., V.12, N.1, 2021, Special Issue, pp.88-103

ON SOME DISPERSION MEASURES FOR FUZZY DATA AND THEIR ROBUSTNESS

PRZEMYSLAW GRZEGORZEWSKI[1,2], KATARZYNA GŁADEK[2]

Abstract. In the present paper we consider measures of dispersion for samples of imprecise data modeled by random fuzzy numbers. Firstly, we suggest a generalization of two well-known classical measures of scale, i.e. the range and interquartile range, for the samples of random fuzzy numbers. Secondly, we examine the robustness of these two measures of dispersion. More precisely, we determine the finite sample breakdown point for each of the introduced measures.

Keywords: breakdown point, dispersion, fuzzy data, outliers, random fuzzy numbers, robustness.

AMS Subject Classification: 62A86, 62G86, 62G35.

1. Introduction

Measures of dispersion (scale) as well as measures of central tendency (location) play a key role in statistics and data analysis. Many tools have been proposed to characterize a sample dispersion, like the range, interquartile range, sample variance, standard deviation, etc. Most of scientific papers on measures of dispersion refer to univariate real-valued data. However, such data do not suffice to model many real-life situations. Besides multidimensional real data or functional data we also need mathematical tools for modeling imprecision that appear quite often in practical applications. Following the seminal paper by Lotfi A. Zadeh [35] fuzzy sets have been recognized as a convenient framework for processing and managing imprecise information.

To formalize a random mechanism generating fuzzy number-valued data within a probabilistic setting Puri and Ralescu [25] introduced random fuzzy numbers (fuzzy random variables). Then, numerous followers have developed various fields of statistics with fuzzy data and fuzzy data analysis. Inferential tools for fuzzy data usually appeared as generalizations of the corresponding concepts applied in classical setting. This is the case of central tendency measures for fuzzy data which have been extensively examined in the literature (see, e.g. [15, 28, 29, 30, 31, 32, 33]). On the other hand, it seems that the variance and standard deviation are the only measures of dispersion for random fuzzy numbers considered in the literature (see, e.g. [4, 6, 23, 26]; for some exceptions see [6]). The lack of measures based on quantiles, like the range or interquartile range, can be explained by the fact that fuzzy numbers are not linearly ordered. However, using a suitable interpretation of the aforementioned measures we can generalize the concepts of the range and the interquartile range so that they could be applied for characterizing the dispersion in the sample of random fuzzy numbers. Such generalization is the first goal of this contribution.

When constructing any estimator one usually undertake an examination of some of its mathematical properties, like unbiasedness, efficiency, consistency, and sufficiency. However, we should also remember that real data often contain outliers which can sometimes involve even substantial distortion in estimation. Thus robustness against outliers should concern both researchers and practitioners. Various approaches for examining robustness of statistical procedures have been proposed in the literature (see [18, 19, 36]). A popular and quite powerful tool for describing

[1]Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland
[2] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
e-mail: pgrzeg@ibspan.waw.pl
Manuscript received January 2021.

the robustness of an estimator is its breakdown point (see [8, 17]). In this context asymptotic research dominate but it seems that the finite sample robustness is much more interesting to practitioners. Thus, the second goal of the present paper is to determine the finite sample breakdown points for the dispersion measures proposed in this contribution.

The paper is organized as follows. In Section 2 we introduce basic notation related to fuzzy sets and fuzzy numbers. Next, in Section 3, we recall some information on random fuzzy numbers and measures of central tendency for fuzzy data. In Section 4 we introduce the generalizations of the range and interquartile range for the samples of random fuzzy numbers. Section 5 contains two theorems on the finite sample breakdown point for the aforementioned dispersion measures.

2. Fuzzy data

The dominant type of data from experiments and statistical observations are real numbers. In the case of imprecise results we need an appropriate mathematical model that will be able to process and manage the available uncertain information. A general framework for handling imprecision was established by the seminal paper by Lotfi A. Zadeh [35] on fuzzy sets.

Let $\mathcal{X}$ be a universe of discourse. A **fuzzy set** A in $\mathcal{X}$ is characterized by its **membership function** $A : \mathcal{X} \to [0, 1]$, which assigns to each object $x \in \mathcal{X}$ a real number in the interval $[0, 1]$, so as $A(x)$ represents the degree of membership of x into A.

The interpretation of $A(x)$ is straightforward: if $A(x) = 1$ then we claim that element $x \in \mathcal{X}$ surely belongs to A, while for $A(x) = 0$ we conclude that x surely does not belong to A. In all other cases, i.e. if $A(x) \in (0, 1)$, we have a partial membership (or partial belongingness) of x into A.

Since the membership function describes completely a corresponding fuzzy set, usually we reduce the notation by identifying a fuzzy set A with its membership function $A(x)$.

Another concept that plays an important role in the theory of fuzzy sets is the so-called **α-cut**. For a fuzzy set A its α-cut, where $\alpha \in [0, 1]$, is defined by

$$A_\alpha = \begin{cases} \{x \in \mathcal{X} : A(x) \geqslant \alpha\} & \text{if } \alpha \in (0, 1], \\ cl\{x \in \mathcal{X} : A(x) > 0\} & \text{if } \alpha = 0, \end{cases} \tag{1}$$

where operator cl stands for the closure. Two α-cuts are of special interest: $A_0 = \operatorname{supp}(A)$ called the **support** and $A_1 = \operatorname{core}(A)$ known as the **core** of a fuzzy number A, respectively. The core indicates all elements of the universe of discourse which surely belong to A, while the support gathers all those elements that possibly belong to A. It is worth underlining that by the decomposition theorem every fuzzy set is completely characterized both by its membership function $A(x)$ and by the family $\{A_\alpha\}_{\alpha \in [0,1]}$ of its all α-cuts.

To generalize real numbers in a way that enables counting of imprecise values and develop both science and its practical applications, fuzzy numbers were introduced by Dubois and Prade [10]. Actually, fuzzy numbers are fuzzy sets in $\mathbb{R}$ which satisfy some additional properties. More precisely, by a **fuzzy number** we consider a mapping $A : \mathbb{R} \to [0, 1]$ such that its α-cut is a nonempty compact interval for each $\alpha \in [0, 1]$.

Many types of fuzzy numbers were proposed in the literature. The most often used fuzzy numbers are **trapezoidal fuzzy numbers** (sometimes called *fuzzy intervals*) with membership functions of the form

$$A(x) = \begin{cases} \frac{x - a_1}{a_2 - a_1} & \text{if } a_1 \leqslant x < a_2, \\ 1 & \text{if } a_2 \leqslant x \leqslant a_3, \\ \frac{a_4 - x}{a_4 - a_3} & \text{if } a_3 < x \leqslant a_4, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $a_1, a_2, a_3, a_4 \in \mathbb{R}$ such that $a_1 \leqslant a_2 \leqslant a_3 \leqslant a_4$. A trapezoidal fuzzy number A is often denoted as $\operatorname{Tra}(a_1, a_2, a_3, a_4)$. Obviously, $a_1 = \inf \operatorname{supp}(A)$, $a_2 = \inf \operatorname{core}(A)$, $a_3 = \sup \operatorname{core}(A)$ and $a_4 = \sup \operatorname{supp}(A)$, which means that each trapezoidal fuzzy numbers is completely described

by its support and core. If $a_2 = a_3$ then A is said to be a **triangular fuzzy number**, while if $a_1 = a_2$ and $a_3 = a_4$ we have a so-called **interval** (or rectangular) fuzzy number.

The families of all fuzzy numbers, trapezoidal fuzzy numbers, triangular fuzzy number and interval fuzzy numbers will be denoted by $\mathbb{F}(\mathbb{R})$, $\mathbb{F}^T(\mathbb{R})$, $\mathbb{F}^{\Delta}(\mathbb{R})$ and $\mathbb{F}^I(\mathbb{R})$, respectively. Obviously, $\mathbb{F}^I(\mathbb{R}) \subset \mathbb{F}^{\Delta}(\mathbb{R}) \subset \mathbb{F}^T(\mathbb{R}) \subset \mathbb{F}(\mathbb{R})$.

To define basic arithmetic operations in $\mathbb{F}(\mathbb{R})$ we use natural α-cut-wise operations on intervals. In particular, the sum of two fuzzy numbers A and B is given by the Minkowski addition of their corresponding α-cuts, i.e.

$$(A + B)_\alpha = \big[\inf A_\alpha + \inf B_\alpha, \sup A_\alpha + \sup B_\alpha\big], \quad \text{for all } \alpha \in [0,1].$$

Similarly, the product of a fuzzy number A by a scalar $\theta \in \mathbb{R}$ is defined by the Minkowski scalar product for intervals, i.e.

$$(\theta \cdot A)_\alpha = \big[\min\{\theta \inf A_\alpha, \theta \sup A_\alpha\}, \max\{\theta \inf A_\alpha, \theta \sup A_\alpha\}\big], \quad \text{for all } \alpha \in [0,1].$$

One may notice that a sum of trapezoidal fuzzy numbers is also a trapezoidal fuzzy number, namely, if $A = \mathrm{Tra}(a_1, a_2, a_3, a_4)$ and $B = \mathrm{Tra}(b_1, b_2, b_3, b_4)$ then

$$A + B = \mathrm{Tra}(a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4). \tag{3}$$

Similarly, the product of a trapezoidal fuzzy number $A = \mathrm{Tra}(a_1, a_2, a_3, a_4)$ by a scalar θ is a trapezoidal fuzzy number

$$\theta \cdot A = \begin{cases} \mathrm{Tra}(\theta a_1, \theta a_2, \theta a_3, \theta a_4) & \text{if } \theta \geqslant 0, \\ \mathrm{Tra}(\theta a_4, \theta a_3, \theta a_2, \theta a_1) & \text{if } \theta < 0. \end{cases} \tag{4}$$

Unfortunately, $\big(\mathbb{F}(\mathbb{R}), +, \cdot\big)$ has not the linear but a semilinear structure since, in general, we have $A + (-1 \cdot A) \neq 1_{\{0\}}$. Consequently, the Minkowski-based difference does not satisfy, in general, the addition/subtraction property that $(A + (-1 \cdot B)) + B = A$. To overcome this problem the so-called Hukuhara difference, defined as follows, was proposed

$$C := A -_H B \quad \text{if and only if} \quad B + C = A.$$

Obviously, the desired properties $A -_H A = 1_{\{0\}}$ or $(A -_H B) + B = A$ are satisfied now. However, the Hukuhara difference is not a completely satisfying resolution of problems with subtraction because it does not always exist.

The aforementioned problems with subtraction in $\mathbb{F}(\mathbb{R})$ do not close the list of inconveniences in statistics with fuzzy numbers. Another critical problem is that the family of fuzzy numbers is not linearly ordered. Hence, in practice, wherever it is possible, it is recommended to avoid subtracting and ranking fuzzy numbers. Obviously, both problems are very inconvenient in applications and stimulate to device solutions based on suitable by-passes.

The problems associated with the lack of a satisfying subtraction operator and univocal ranking could be overcome in statistical reasoning by developing an alternative approach based on suitable metrics in $\mathbb{F}(\mathbb{R})$ (see, e.g. [4]). One can define various metrics in $\mathbb{F}(\mathbb{R})$ but in this contribution we make use of the following ones.

Definition 2.1 ([7]). *Let $A, B \in \mathbb{F}(\mathbb{R})$. The following mapping $\rho_1 : \mathbb{F}(\mathbb{R}) \times \mathbb{F}(\mathbb{R}) \to [0, +\infty)$*

$$\rho_1(A, B) = \frac{1}{2} \int_{(0,1]} \Big(\big|\inf A_\alpha - \inf B_\alpha\big| + \big|\sup A_\alpha - \sup B_\alpha\big| \Big)\, d\alpha$$

*is called the ρ_1-**norm distance** between A and B, while the mapping $\rho_2 : \mathbb{F}(\mathbb{R}) \times \mathbb{F}(\mathbb{R}) \to [0, +\infty)$ defined as*

$$\rho_2(A, B) = \sqrt{\frac{1}{2} \int_{(0,1]} \Big(\big[\inf \inf A_\alpha - \inf B_\alpha\big]^2 + \big[\sup A_\alpha - \sup B_\alpha\big]^2 \Big)\, d\alpha},$$

*is called the ρ_2-**norm distance** between fuzzy numbers A and B.*

Definition 2.2 ([28]). *Let $A, B \in \mathbb{F}(\mathbb{R})$. The mapping $D_1 : \mathbb{F}(\mathbb{R}) \times \mathbb{F}(\mathbb{R}) \to [0, +\infty)$ defined by*

$$D_1(A,B) = \big|\mathrm{wabl} - \mathrm{wabl}(B)\big| + \frac{1}{2}\int\limits_{[0,1]} \Big(\big|\mathrm{ldev}_A(\alpha) - \mathrm{ldev}_B(\alpha)\big| + \big|\mathrm{rdev}_A(\alpha) - \mathrm{rdev}_B(\alpha)\big|\Big)\, d\alpha,$$

where

$$\begin{aligned} \mathrm{wabl}(A) &= \int_{[0,1]} \mathrm{mid}\, A_\alpha \, d\alpha, \\ \mathrm{ldev}_A(\alpha) &= \mathrm{wabl}(A) - \inf A_\alpha, \\ \mathrm{rdev}_A(\alpha) &= \sup A_\alpha - \mathrm{wabl}(A), \end{aligned}$$

with $\mathrm{mid}\, A_\alpha$ *denoting the center point (mid-point) of* A_α*, is called the* ***wabl/ldev/rdev-based*** L^1 *distance between fuzzy numbers A and B (further on called simply* D_1***-distance**).*

It is seen that metrics ρ_1 and D_1 are the L^1-type on $\mathbb{F}(\mathbb{R})$, while ρ_2 is the L^2-type metric. All three metric spaces $(\mathbb{F}(\mathbb{R}), \rho_1)$, $(\mathbb{F}(\mathbb{R}), \rho_2)$ and $(\mathbb{F}(\mathbb{R}), D_1)$ are separable. Moreover, metric spaces $(\mathbb{F}(\mathbb{R}), \rho_1)$ and $(\mathbb{F}(\mathbb{R}), \rho_2)$, through the support function of fuzzy sets and aforementioned arithmetic, can be isometrically embedded onto a convex cone of the Banach space of the L^1-type real-valued functions defined on $[0,1] \times \{-1, 1\}$ with the functional arithmetic and the distance induced by certain norms (see [6, 7]). On the other hand, the metric space $(\mathbb{F}(\mathbb{R}), D_1)$ can be isometrically embedded onto a convex cone of the Hilbert space of the L^2-type real-valued functions defined on $[0,1] \times \{-1, 1\}$ with the functional arithmetic and the distance induced by a certain norm (see [7]).

For more details on fuzzy numbers, their types, characteristics, and approximations we refer the reader to [2].

3. Random fuzzy numbers

Suppose that the experiment results generate a random sample of imprecise data described by fuzzy numbers. To formalize a mathematical model which allows to grasp both aspects of uncertainty that appear in such data, i.e. randomness associated with data generation and fuzziness connected with their imprecision, Puri and Ralescu [25] introduced the notion of a **fuzzy random variable**, called also a **random fuzzy number**.

Definition 3.1. *Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathbb{F}(\mathbb{R})$ is said to be a* ***random fuzzy number*** *if for all $\alpha \in [0,1]$ the α-level function $X_\alpha(\omega) = \big(X(\omega)\big)_\alpha$ is a compact random interval.*

Equivalently, $X : \Omega \to \mathbb{F}(\mathbb{R})$ is a random fuzzy number if for all $\alpha \in [0,1]$ the real-valued mappings $\inf X_\alpha$ and $\sup X_\alpha$ are usual real-valued random variables. In other words, X is a random fuzzy number if and only if X is a Borel measurable function w.r.t. the Borel σ-field generated on $\mathbb{F}(\mathbb{R})$ by the topology induced by metrics like those in Definition 2.1 or Definition 2.2. Due to Borel-measurability we may properly refer to the distribution induced by a random fuzzy number, the stochastic independence of random fuzzy numbers, etc.

Both descriptive and inferential statistics utilize various summary statistics of a sample. The most widely applied are measures of location, especially measures of central tendency such as the mean, median and so on. Obviously the same happens in statistics with fuzzy data where some natural counterparts of those characteristics of location has been defined.

Definition 3.2 ([25]). *Let $(\Omega, \mathcal{A}, \mathcal{P})$ denote a probabilistic space and let $X : \Omega \to \mathbb{F}(\mathbb{R})$ be a random fuzzy number such that $\mathbb{E}(\max\{|\inf X_0|, |\sup X_0|\}) < \infty$. The* ***Aumann mean*** *of X is a fuzzy number $\widetilde{\mathbb{E}}(X) \in \mathbb{F}(\mathbb{R})$ such that*

$$\big(\widetilde{\mathbb{E}}(X)\big)_\alpha = \big[\mathbb{E}(\inf X_\alpha),\, \mathbb{E}(\sup X_\alpha)\big], \quad \forall \alpha \in [0,1],$$

where $\mathbb{E}$ *stands for the expected value of a real-valued random variable.*

If $\mathbb{X}_n = (X_1, ..., X_n)$ is a sample of random fuzzy numbers then the most common location measure of that sample is the sample mean $\overline{\mathbb{X}}_n$ defined by its α-cuts as follows

$$\big(\overline{\mathbb{X}}_n\big)_\alpha = \Big(\frac{1}{n}\cdot(X_1 + \ldots + X_n)\Big)_\alpha = \Big[\frac{1}{n}\sum_{i=1}^{n}\inf(X_i)_\alpha\,,\ \frac{1}{n}\sum_{i=1}^{n}\sup(X_i)_\alpha\Big].$$

where $\alpha \in [0,1]$.

In a similar way one may define a fuzzy analogue of the population median and the sample median for fuzzy random variables. Actually, we may consider several variants of those definitions depending on the underlying distance between fuzzy numbers.

Definition 3.3 ([30])**.** *The* ρ_1***-median*** *of a random fuzzy number* X *is a fuzzy number* $\widetilde{\mathrm{Me}}(X)$ *such that*

$$\big(\widetilde{\mathrm{Me}}(X)\big)_\alpha = \big[\mathrm{Me}(\inf X_\alpha),\, \mathrm{Me}(\sup X_\alpha)\big], \quad \forall \alpha \in [0,1],$$

where Me *stands for the usual median of a real-valued random variable.*

Given a sample of random fuzzy numbers $\mathbb{X}_n = (X_1, ..., X_n)$ the sample ρ_1-median is a fuzzy number $\widehat{\widetilde{\mathrm{Me}}}(\mathbb{X}_n)$ defined by its α-cuts, $\alpha \in [0,1]$, as follows

$$\big(\widehat{\widetilde{\mathrm{Me}}}(\mathbb{X}_n)\big)_\alpha = \Big[\mathrm{Me}\big(\inf(X_1)_\alpha, \ldots, \inf(X_n)_\alpha\big), \mathrm{Me}\big(\sup(X_1)_\alpha, \ldots, \sup(X_n)_\alpha\big)\Big].$$

Definition 3.4 ([28])**.** *The* D_1***-median*** *of a random fuzzy number* X *is a fuzzy number* $\widetilde{\mathrm{Me}}(X)$ *such that*

$$\big(\mathrm{M}(X)\big)_\alpha = \Big[\mathrm{Me}\big(\mathrm{wabl}(X)\big) - \mathrm{Me}\big(\mathrm{ldev}_X(\alpha)\big), \mathrm{Me}\big(\mathrm{wabl}(X)\big) + \mathrm{Me}\big(\mathrm{rdev}_X(\alpha)\big)\Big], \quad \forall \alpha \in [0,1],$$

where Me *stands, as before, for the usual median of a real-valued random variable.*

The sample D_1-median is a fuzzy number $\widehat{\mathrm{M}}(\mathbb{X}_n)$ defined by its α-cuts, $\alpha \in [0,1]$, as follows

$$\begin{aligned}\big(\widehat{\mathrm{M}}(\mathbb{X}_n)\big)_\alpha = \Big[&\mathrm{Me}\big(\mathrm{wabl}(X_1), \ldots, \mathrm{wabl}(X_n)\big) - \mathrm{Me}\big(\mathrm{ldev}_{X_1}(\alpha), \ldots, \mathrm{ldev}_{X_n}(\alpha)\big),\\ &\mathrm{Me}\big(\mathrm{wabl}(X_1), \ldots, \mathrm{wabl}(X_n)\big) + \mathrm{Me}\big(\mathrm{rdev}_{X_1}(\alpha), \ldots, \mathrm{rdev}_{X_n}(\alpha)\big)\Big].\end{aligned}$$

Next lemma gives the values of $\mathrm{wabl}\big(\widehat{\mathrm{M}}(\mathbb{X}_n)\big)$, $\mathrm{ldev}_{\widehat{\mathrm{M}}(\mathbb{X}_n)}(\alpha)$ and $\mathrm{rdev}_{\widehat{\mathrm{M}}(\mathbb{X}_n)}(\alpha)$ which are utilized later on (e.g. in the proof of Theorem 5.2).

Lemma 3.1. *Let* $\mathbb{X}_n = (X_1, \ldots, X_n)$ *be a sample of random fuzzy numbers. Then*

$$\begin{aligned}\mathrm{wabl}\big(\widehat{\mathrm{M}}(\mathbb{X}_n)\big) &= \mathrm{Me}\left(\mathrm{wabl}\,(X_1), \ldots, \mathrm{wabl}\,(X_n)\right)\\ &\quad + \frac{1}{2}\int_{[0,1]} \Big[\mathrm{Me}\big(\mathrm{rdev}_{X_1}(\alpha), \ldots, \mathrm{rdev}_{X_n}(\alpha)\big) - \mathrm{Me}\big(\mathrm{ldev}_{X_1}(\alpha), \ldots, \mathrm{ldev}_{X_n}(\alpha)\big)\Big]\, d\alpha,\\ \mathrm{ldev}_{\widehat{\mathrm{M}}(\mathbb{X}_n)}(\alpha) &= \mathrm{Me}\big(\mathrm{ldev}_{X_1}(\alpha), \ldots, \mathrm{ldev}_{X_n}(\alpha)\big)\\ &\quad + \frac{1}{2}\int_{[0,1]} \Big[\mathrm{Me}\big(\mathrm{rdev}_{X_1}(\alpha), \ldots, \mathrm{rdev}_{X_n}(\alpha)\big) - \mathrm{Me}\big(\mathrm{ldev}_{X_1}(\alpha), \ldots, \mathrm{ldev}_{X_n}(\alpha)\big)\Big] d\alpha,\\ \mathrm{rdev}_{\widehat{\mathrm{M}}(\mathbb{X}_n)}(\alpha) &= \mathrm{Me}\big(\mathrm{rdev}_{X_1}(\alpha), \ldots, \mathrm{rdev}_{X_n}(\alpha)\big)\\ &\quad - \frac{1}{2}\int_{[0,1]} \Big[\mathrm{Me}\big(\mathrm{rdev}_{X_1}(\alpha), \ldots, \mathrm{rdev}_{X_n}(\alpha)\big) - \mathrm{Me}\big(\mathrm{ldev}_{X_1}(\alpha), \ldots, \mathrm{ldev}_{X_n}(\alpha)\big)\Big] d\alpha.\end{aligned}$$

The proof of the lemma reduces to some straightforward transformations so it is left to the reader.

4. Dispersion measures for fuzzy data

Although measures of central tendency dominate in statistical applications it is well known that usually they are not sufficient to describe a sample. For example, two samples with a similar mean may differ significantly in dispersion which do not allow to conclude that they come from the same distribution. Similarly, usually is not enough to determine if a manufacturing or business process is in a state of control using the $\overline{x}$ chart only. In statistical process control (SPC) it is always recommended to monitor a process simultaneously by a pair of $\overline{x}$–R or $\overline{x}$–S charts. Indeed, if the sample variability itself is not in statistical control, then the entire process cannot be considered to be in control regardless of what the $\overline{x}$ chart indicates.

Before starting the discussion on particular measures of dispersion for fuzzy data let us consider basic requirements which each such measure should satisfy. Recently, Kołacz and Grzegorzewski [21] proposed an axiomatic definition of a measure of dispersion for a sample in $\mathbb{R}^n$. Let us adopt their definition for fuzzy data.

Definition 4.1. *A function $\Delta : \bigcup_{n=1}^{\infty}(\mathbb{F}(\mathbb{R}))^n \to [0,\infty)$ is called a **measure of dispersion** if Δ is a non-identically zero function satisfying the following axioms for any $(X_1,\ldots,X_n) \in (\mathbb{F}(\mathbb{R}))^n$:*

(A1) $\Delta(X_1,\ldots,X_n) = 0$, *if* $X_1 = \ldots = X_n$.

(A2) Δ *is symmetric, i.e.* $\Delta(X_{\pi(1)},\ldots,X_{\pi(n)}) = \Delta(X_1,\ldots,X_n)$ *for any permutation* $\pi : \{1,\ldots,n\} \to \{1,\ldots,n\}$.

(A3) Δ *is translation invariant, i.e.* $\Delta(X_1 + t,\ldots,X_n + t) = \Delta(X_1,\ldots,X_n) \quad \forall t \in \mathbb{F}(\mathbb{R})$.

Sometimes one more axiom is also considered:

(A4) There exists a function $\zeta : \mathbb{R} \to [0,\infty)$ *such that* $\Delta(aX_1,\ldots,aX_n) = \zeta(a)\Delta(X_1,...,X_n)$ $\forall a \in \mathbb{R}^+$.

4.1. The variance of fuzzy data.

The most popular dispersion measure applied for fuzzy data is the so-called Fréchet variance defined as follows.

Definition 4.2 ([11])**.** *Let $(\Omega,\mathcal{A},\mathcal{P})$ be a probabilistic space. The **Fréchet variance** of a random fuzzy number $X : \Omega \to \mathbb{F}(\mathbb{R})$ is given by*

$$\widetilde{\sigma}_X^2 = \mathbb{E}\Big(\rho_2^2\big[X,\widetilde{\mathbb{E}}(X)\big]\Big),$$

where ρ_2 is the distance specified in Definition 2.1.

Having a sample of fuzzy random numbers $\mathbb{X}_n = (X_1,...,X_n)$, the sample Fréchet variance is given by

$$\widetilde{S}^2[\mathbb{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\rho_2^2\big(X_i,\overline{\mathbb{X}}_n\big).$$

As it is seen, the Fréchet variance for fuzzy random numbers assumes a non-negative real value. It is worth noting that Kruse and Meyer [22] suggested a fuzzy sample variance for fuzzy data. However, their definition refers to the *epistemic view* on fuzzy data, contrary to the *ontic view* discussed in this paper. For the overview on both views on fuzzy data we refer the reader to [5].

4.2. The range of a fuzzy sample.

As it is known, the range of a real-valued sample $(\xi_1,\ldots,\xi_n)$ is defined as the difference between the largest and the smallest observation, i.e. $\mathrm{R} = \xi_{n:n} - \xi_{1:n}$, where $\xi_{i:n}$ denotes the i-th biggest observation in the sample (the i-th order statistic). The aforementioned definition is not suitable for the straightforward generalization of the range into the fuzzy domain since there is no natural linear ordering in the family of fuzzy numbers. However, one may easily notice that the range of a real-valued sample $(\xi_1,\ldots,\xi_n)$

might be expressed equivalently as the biggest distance between any two observations in the sample, i.e.

$$\mathrm{R} = \max\left\{\xi_i - \xi_j : i, j = 1, \ldots, n\right\}. \tag{5}$$

This remark serves as an inspiration for the proposed definition of the range of a sample consisting of fuzzy observations.

Definition 4.3. *Let d denote a distance in $\mathbb{F}(\mathbb{R})$. Then the* ***d-range*** *of a sample $\mathbb{X}_n = (X_1, ..., X_n)$ of random fuzzy numbers is given by*

$$\widetilde{\mathrm{R}}_d[\mathbb{X}_n] = \max\{d(X_i, X_j) : i, j = 1, \ldots, n\}.$$

Keeping in mind Definition 4.1 one can easily prove the following lemma.

Lemma 4.1. *The d-range is a dispersion measure.*

A distance is a two-argument function measuring how much any two points of the given space are separated. Martín and Mayor [24] extended this definition so it might be applied to a collection of more than two elements and called it a *multidistance*, defined as follows.

Definition 4.4. *A function* $\mathrm{MultD} : \bigcup_{n\geqslant 1}(\mathbb{R}^p)^n \to [0, \infty)$ *is called a* ***multidistance*** *if it satisfies the following conditions for all $\mathbf{x}_1, ..., \mathbf{x}_n, \mathbf{y} \in \mathbb{R}^p$:*

(md1) $\mathrm{MultD}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = 0$ *if and only if* $\mathbf{x}_1 = \ldots = \mathbf{x}_n$.
(md2) $\mathrm{MultD}(\mathbf{x}_{\pi(1)}, \ldots, \mathbf{x}_{\pi(n)}) = \mathrm{MultD}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ *for any permutation π of $\{1, \ldots, n\}$.*
(md3) $\mathrm{MultD}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \leqslant \mathrm{MultD}(\mathbf{x}_1, \mathbf{y}) + \ldots + \mathrm{MultD}(\mathbf{x}_n, \mathbf{y})$.

Obviously, we may easily generalize Definition 4.4 into fuzzy domain substituting observations $\mathbb{R}^p$ by fuzzy data from $\mathbb{F}(\mathbb{R})$. Then, the following proposition could be proved immediately.

Proposition 4.1. *The d-range is a multidistance.*

Proposition 4.1 indicates another (in some sense inverse) way for defining the range of a sample of fuzzy observations.

Definition 4.5. *Let* MultD *denote a multidistance in $\mathbb{F}(\mathbb{R})$. Then the* ***range based on the multidistance*** MultD *of a sample $\mathbb{X}_n = (X_1, ..., X_n)$ of random fuzzy numbers is given by*

$$\widetilde{\mathrm{R}}_{\mathrm{MultD}}[\mathbb{X}_n] = \mathrm{MultD}(X_1, \ldots, X_n). \tag{6}$$

Thus it is clear that by a suitable choice of a multidistance (e.g. the diameter of the smallest ball containing all observations [24]), we obtain alternative definitions of the range.

Let us complete this section with a remarks that substituting random fuzzy numbers in Definition 4.3 or 4.5 by random intervals we obtain the corresponding concepts introduced in [16] for the interval data.

4.3. **The interquartile range of a fuzzy sample.** Because of high sensitivity to outliers the range is relatively rare applied in practice. To avoid the influence of possible outliers on the overall measure one may use the interquartile range $\mathrm{IQR} = Q_3 - Q_1$ defined as a difference between the upper Q_3 and lower quartile Q_1, i.e. the 75th and 25th percentiles, respectively. Actually, the aforementioned definition of the interquartile range may appear not so straightforward in calculations since there is no unique way to determine percentiles in a sample. Consequently, if $\xi_1, \ldots, \xi_n$ denotes a real-valued sample we may consider the following general formula for the interquartile range

$$\mathrm{IQR} = (1 - \gamma)\left(\xi_{k:n} - \xi_{l:n}\right) + \gamma\left(\xi_{k+1:n} - \xi_{l+1:n}\right), \tag{7}$$

where $k = \lfloor 0.75n + m \rfloor$ and $l = \lfloor 0.25n + m \rfloor$, while m and γ are some constants which depend on a particular method for determining percentiles (see, e.g., [20, 21]).

Here one may immediately conclude that neither the first simple definition of IQR nor the second, more sophistical one qualifies for the generalization into fuzzy domain, because of the

problem with ranking fuzzy numbers. However, this remark does not mean that we are completely helpless. As in the case of the range we should start from the deeper insight into the nature of the considered dispersion measure. Here one may notice that the interquartile range for the real-valued sample is equal to the width of the interval containing about 50% of the middle observations Such a procedure allows to lose about 50% of questionable observations which are relatively far from the center of the sample. Since in a sample of real-valued data outliers are either too big or too small observations, in practice we omit about 25% of the smallest and about 25% of the biggest observations.

One may be curious why do we mention "about 50%" or "about 25%" of observations instead of saying strictly 50% or 25%, respectively. It is so because depending on the sample size n and a particular method for determining percentiles the interquartile range (7) may contain $50\% \pm 1$ or 2 of the middle observations.

Now, before proposing a generalization of the interquartile range for fuzzy data we have to resolve two issues: *How to choose the center of a fuzzy sample?* and *Which observations are worth to omit as possible outliers?*

As for the first point, we have discussed several measures of central tendency for a fuzzy sample in Section 3. Each of those measures could be considered as a "center"of the sample. However, to avoid the influence of possible outliers we take the sample median as the "typical" observation.

Turning now to the second point it becomes obvious that the notion of outlier in fuzzy context means something different than in the real-valued case. Here, being too big or too small compared to most observations in a sample does not exhaust the meaning of being an outlier. It is so because fuzzy numbers are not characterized only by their location. Indeed, membership functions of two fuzzy numbers with identical support may differ vastly. Actually, if two persons consider a mathematical model of a linguistic variable assuming, e.g. "about 5", we may obtain not necessarily identical fuzzy numbers: obviously, we would expect to find 5 in the support of both fuzzy numbers but one might be, e.g., triangular, while the other parabolic and so on. Even if they are both of the same type, e.g., trapezoidal, they may differ in support or core.

Therefore, we suggest to consider as possible outliers those fuzzy observations whose distance from the sample center is large. Hence, keeping in mind prior arrangements, we will omit observations far enough from the sample median.

Let $\mathbb{X}_n = (X_1, \ldots, X_n)$ denote a sample of random fuzzy numbers and let $\widetilde{\mathrm{Med}}[\mathbb{X}]$ be a sample median (one of those defined in Section 3). Moreover, let d denote a distance in $\mathbb{F}(\mathbb{R})$ (one of those defined in Section 2). For a fixed $\widetilde{\mathrm{Med}}[\mathbb{X}]$ and d we compute the distance between each fuzzy observation X_i and the median, i.e. obtain the following sequence $(\zeta_1, \ldots, \zeta_n)$, where $\zeta_i = d(X_i, \widetilde{\mathrm{Med}}[\mathbb{X}])$. Next, we order this sequence $(\zeta_1, \ldots, \zeta_n)$, so we obtain $(\zeta_{1:n}, \ldots, \zeta_{n:n})$. Let $\zeta^* = \zeta_{k:n}$, where $k = \lfloor \frac{n}{2} \rfloor$ denotes the k-th biggest distance from the sample median $\widetilde{\mathrm{Med}}[\mathbb{X}]$.

Let us define the following subset $\mathbb{X}_n^{1/2}$ of the original sample $\mathbb{X}_n$

$$\mathbb{X}_n^{1/2} = \left\{ X_i \in \mathbb{X} : d(X_i, \widetilde{\mathrm{Med}}[\mathbb{X}_n]) \leqslant \zeta^* \right\}. \tag{8}$$

One may notice that $\mathbb{X}_{1/2}$ contains "about 50% of the central" observations, i.e. only such fuzzy observations which do not differ too much from the sample median. Now we can find the biggest distance between observations that remained in $\mathbb{X}_n^{1/2}$. This very distance would be considered as the interquartile range of a fuzzy sample. More formally, we obtain the following definition.

Definition 4.6. *Let d denote a distance in $\mathbb{F}(\mathbb{R})$. Then the **d-interquartile range** of a sample $\mathbb{X}_n = (X_1, ..., X_n)$ of random fuzzy numbers is given by*

$$\widetilde{\mathrm{IQR}}_d[\mathbb{X}_n] = \max \left\{ d(X_i, X_j) : X_i, X_j \in \mathbb{X}_n^{1/2} \right\}. \tag{9}$$

One can proof the following theorem.

Lemma 4.2. *The d_θ-interquartile range is a dispersion measure.*

We may also define the interquartile range with respect to a multidistance.

Definition 4.7. *Let* MultD *denote a multidistance in* $\mathbb{F}(\mathbb{R})$. *Then the* ***interquartile range based on the multidistance*** MultD *of a sample* $\mathbb{X}_n = (X_1, ..., X_n)$ *of random fuzzy numbers is given by*

$$\widetilde{\mathrm{IQR}}_{\mathrm{MultD}}[\mathbb{X}] = \mathrm{MultD}(\mathbb{X}_n^{1/2}). \tag{10}$$

Please, notice that substituting random fuzzy numbers in Definition 4.6 or 4.7 by random intervals we obtain the corresponding concepts introduced in [16] for the interval data.

5. Robustness of the dispersion measures

We can consider different aspects of robustness in statistics. However, at the starting point one should answer two fundamental questions: *Robustness against what?* and *Robustness with respect to what?* In this section we will remain in the traditional current of research and examine robustness of the dispersion measures introduced in the previous sections against outliers. Various tools have been proposed in quantitative robustness. On of the most appreciated one is the **breakdown point** originally proposed by Hampel [17] and generalized by Huber [19]. The concept of the breakdown point restricted to finite samples, the so-called **finite sample breakdown point** (or fsbp for short), was suggested by Donoho [8] and Donoho and Huber [9].

Intuitively, the finite sample breakdown point of an estimator is the proportion of ,,incorrect" observations in a sample an estimator can handle before giving a wrong result. For instance, the breakdown point of a location measure is the proportion of observations arbitrarily far from the center (i.e. large or small) that this measure can handle before giving result arbitrarily far from the true sample location (which is sometimes called the **explosion**). Obviously, the higher breakdown point, the more robust is the estimator under study.

The finite sample breakdown point for measures of dispersion is defined as the minimum proportion of observations in a sample which should be perturbed to let the measure get values either arbitrary large (explosion) or equal to zero (sometimes called the *implosion*). Therefore, we have to consider two situations: the first one with a sample containing outliers making the estimator overestimate the true dispersion up to infinity, and the second one with a sample containing inliers leading to underestimation of the true dispersion to zero. A formal definition of the finite sample breakdown point for measures of dispersion (see [9]) adapted to fuzzy observations is given below.

Definition 5.1. *Let* $\mathbb{X}_n = (X_1, ..., X_n)$ *be a sample of random fuzzy numbers and let* $x_n = (x_1, ..., x_n)$ *denote a realization of the sample* $\mathbb{X}_n$. *Then the* ***finite sample breakdown point of a dispersion measure*** T *for the sample* $\curvearrowleft_n$ *is defined by*

$$\mathrm{fsbp}^*(T, \curvearrowleft_n) = \min\left\{\mathrm{fsbp}^+(T, \curvearrowleft_n), \mathrm{fsbp}^-(T, \curvearrowleft_n)\right\},$$

where

$$\mathrm{fsbp}^+(T, x_n) = \min\left\{\frac{k}{n} : \sup_{y_{n,k}} T(y_{n,k}) = \infty\right\} \tag{11}$$

$$\mathrm{fsbp}^-(T, x_n) = \min\left\{\frac{k}{n} : \inf_{y_{n,k}} T(y_{n,k}) = 0\right\}, \tag{12}$$

with $y_{n,k}$ *denoting a sample obtained by replacing any* k *observations of the sample* x_n *by arbitrary values. The quantities* $fsbp^+$ *i* $fsbp^-$ *are called the* ***explosion breakdown point*** *and the* ***implosion breakdown point****, respectively.*

It is worth emphasizing that when searching for measures of dispersion, the breakdown point turns out to have a considerable practical importance. In some cases the breakdown point is more important than the efficiency of any corresponding estimator.

For the samples containing fuzzy observations the breakdown points for the standard deviation, the average distance deviation about the mean and the median distance deviation about the median were determined in [6]. Below we examine the finite sample breakdown points for the range and for the interquartile range of random fuzzy numbers introduced in Section 4.2 and Section 4.3, respectively.

Theorem 5.1. *Let $\mathbb{X}_n = (X_1, \ldots, X_n)$ be a sample of random fuzzy numbers and let $x_n = (x_1, \ldots, x_n)$ denote its realization such that there are no two identical fuzzy numbers in x_n. Let d denote a distance in $\mathbb{F}(\mathbb{R})$ such that $d \in \{\rho_1, \rho_2, D_1\}$. Then*

$$\mathrm{fsbp}^+(\widetilde{R}_d, x_n) = \frac{1}{n} \quad \textit{and} \quad \mathrm{fsbp}^-(\widetilde{R}_d, x_n) = \frac{n-1}{n}.$$

Consequently, the finite sample breakdown of $\widetilde{R}_d$ equals

$$\mathrm{fsbp}^*(\widetilde{R}_d, x_n) = \frac{1}{n}.$$

Proof. Suppose $\mathbb{X}_n = (X_1, \ldots, X_n)$ is a sample of random fuzzy numbers and let $x_n = (x_1, \ldots, x_n)$ denote its realization having no identical fuzzy numbers in x_n. The proof is split into three steps.

Step 1: We begin showing that $\mathrm{fsbp}^-(\widetilde{R}_d, x_n) \leqslant \frac{n-1}{n}$.

We construct a sample $y_{n,n-1}$ such that $y_1 = x_1$ and $y_2, \ldots, y_n = x_1$, i.e. $y_{n,n-1}$ originated from x_n by substituting $n-1$ observations with x_1. Notice, that

$$\widetilde{R}_d(y_{n,n-1}) = \widetilde{R}_d(y_1, ..., y_n) = \widetilde{R}_d(x_1, ..., x_1) = \max\{d(x_i, x_j) : i, j = 1\} = 0.$$

Thus $\inf_{z_{n,n-1}} \widetilde{R}_d(z_{n,n-1}) = 0$, where $z_{n,n-1}$ any sample originated from x_n by replacing $n-1$ observations, which implies $\mathrm{fsbp}^-(\widetilde{R}_d, x_n) \leqslant \frac{n-1}{n}$.

Step 2: We show that $\mathrm{fsbp}^-(\widetilde{R}_d, x_n) \geqslant \frac{n-1}{n}$.

Let $y_{n,k}$ be a sample obtained from x_n by substituting $k < n-1$ observations. Then there exist at least two observations $x^*, x^{**} \in \{x_1, ..., x_n\}$ such that $x^*, x^{**} \in y_{n,k}$. Let $\delta = d(x^*, x^{**}) > 0$ since we have assumed no identical observations in x_n. Hence

$$\delta = d(x^*, x^{**}) \leqslant \max\{d(y_i, y_j) : i, j = 1, \ldots, n\} = \widetilde{R}_d(y_{n,k}).$$

So $\inf_{z_{n,k}} \widetilde{R}_d(z_{n,k}) \geqslant \delta > 0$ for $k < n-1$, and therefore $\mathrm{fsbp}^-(\widetilde{R}_d, x_n) \geqslant \frac{n-1}{n}$.

To sum up both Step 1 and Step 2 we obtain $\mathrm{fsbp}^-(\widetilde{R}_d, x_n) = \frac{n-1}{n}$.

Step 3: We will show that $\mathrm{fsbp}^+(\widetilde{R}_d, x_n) = \frac{1}{n}$.

We construct a sample $y_{n,1}$ by replacing a single observation in x_n. Let $y_1 \in \mathbb{F}(\mathbb{R})$ such that $(y_1)_\alpha = [\inf (x_2)_\alpha, \sup (x_2)_\alpha + 2L]$, where $L \in \mathbb{R}$ and $L > 0$. Other observations, i.e. $y_2 = x_2, \ldots, y_n = x_n$, remain without changing. Now let us compute $d(y_1, y_2)$ for various distances, i.e. $d \in \{\rho_1, \rho_2, D_1\}$. We obtain

$$\begin{aligned}
\rho_1(y_1, y_2) &= \frac{1}{2} \int_{(0,1]} \left(|\inf (y_1)_\alpha - \inf (y_2)_\alpha| + |\sup (y_1)_\alpha - \sup (y_2)_\alpha|\right) d\alpha \\
&= \frac{1}{2} \int_{(0,1]} |L| d\alpha = L, \\
\rho_2(y_1, y_2) &= \sqrt{\frac{1}{2} \int_{(0,1]} \left([\inf (y_1)_\alpha - \inf (y_2)_\alpha]^2 + [\sup (y_1)_\alpha - \sup (y_2)_\alpha]^2\right) d\alpha} \\
&= \sqrt{\frac{1}{2} \int_{(0,1]} 4L^2 d\alpha} = L\sqrt{2},
\end{aligned}$$

W also have

$$\begin{aligned}\operatorname{wabl}(y_1) &= \int\limits_{[0,1]} \operatorname{mid}(y_1)_\alpha \, d\alpha = \int\limits_{[0,1]} \left(\frac{\inf(x_2)_\alpha + \sup(x_2)_\alpha}{2} + L \right) d\alpha \\ &= L + \operatorname{wabl}(x_2) = L + \operatorname{wabl}(y_2), \\ \operatorname{ldev}_{y_1}(\alpha) &= \operatorname{wabl}(y_1) - \inf(y_1)_\alpha = L + \operatorname{wabl}(y_2) - \inf(x_2)_\alpha \\ &= L + \operatorname{ldev}_{y_2}(\alpha), \\ \operatorname{rdev}_{y_1}(\alpha) &= \sup(y_1)_\alpha - (L + \operatorname{wabl}(y_2)) = \sup(x_2)_\alpha + 2L - L - \operatorname{wabl}(y_2) \\ &= L + \operatorname{rdev}_{y_2}(\alpha),\end{aligned}$$

which leads to

$$\begin{aligned}D_1(y_1, y_2) &= \big|\operatorname{wabl}(y_1) - \operatorname{wabl}(y_2)\big| \\ &\quad + \frac{1}{2}\int_0^1 \Big(\big|\operatorname{ldev}_{y_1}(\alpha) - \operatorname{ldev}_{y_2}(\alpha)\big| + \big|\operatorname{rdev}_{y_1}(\alpha) - \operatorname{rdev}_{y_2}(\alpha)\big| \Big) d\alpha \\ &= |L| + \frac{1}{2}\int_0^1 (|L| + |L|) d\alpha = 2L.\end{aligned}$$

Finally,

$$\widetilde{R}_d(y_{n,1}) = \max\{d(y_i, y_j)_{i,j=1,\ldots,n}\} \geq d(y_1, y_2) \geq \min\{L, L\sqrt{2}, 2L\}.$$

Letting $L \to \infty$ we obtain $\sup_{\widetilde{z}_{n,1}} \widetilde{R}_d(\widetilde{z}_{n,1}) = \infty$ for any sample $\widetilde{z}_{n,1}$ originated from x_n by replacing a single observation. Therefore, $\operatorname{fsbp}^+(\widetilde{R}_d, x_n) = \frac{1}{n}$, which completes the proof. □

Theorem 5.2 shows that the range of a sample of fuzzy numbers is not robust against outliers since its finite sample breakdown point assumes $\frac{1}{n}$ which is the lowest possible value. On the other hand it is not disappointing since the same happens for the real-valued data. Moreover, this result is obtained for various distances between fuzzy numbers. Next theorem examines the robustness of the interquartile range. However, contrary to Theorem 5.2, we restrict our attention to metric D_1.

Theorem 5.2. *Let $\mathbb{X}_n = (X_1, \ldots, X_n)$ be a sample of random fuzzy numbers and let $x_n = (x_1, \ldots, x_n)$ denote its realization such that there are no two identical fuzzy numbers in x_n. Then*

$$\operatorname{fsbp}^+(\widetilde{IQR}_{D_1}, x_n) = \frac{\lceil \frac{n}{2} \rceil}{n} \quad \textit{and} \quad \operatorname{fsbp}^-(\widetilde{IQR}_{D_1}, x_n) = \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}.$$

Consequently, the finite sample breakdown of $\widetilde{IQR}_{D_1}$ equals

$$\operatorname{fsbp}^*(\widetilde{IQR}_{D_1}, x_n) = \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}.$$

Proof. Suppose $\mathbb{X}_n = (X_1, \ldots, X_n)$ is a sample of random fuzzy numbers and let $x_n = (x_1, \ldots, x_n)$ denote its realization having no identical fuzzy numbers in x_n. The proof is split into four steps.

Step 1: Firstly, we show that $\operatorname{fsbp}^-(\widetilde{IQR}_{D_1}, x_n) \leqslant \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}$.

We construct a sample $y_{n,\lfloor \frac{n}{2} \rfloor - 1}$ by replacing $\lfloor \frac{n}{2} \rfloor - 1$ observations in the original sample x_n. More precisely, we substitute observations whose distance from the median $\widehat{M}[x_n]$ is equal to $\zeta_2, \ldots, \zeta_{\lfloor \frac{n}{2} \rfloor}$ with the observation that is located at a distance ζ_1 from $\widehat{M}[x_n]$. Further on we will denote this very observation by x_{ζ_1}. As a result $\lfloor \frac{n}{2} \rfloor$ become identical and they all are located at the minimal distance from $\widehat{M}[x_n]$ among all observations in $y_1, \ldots, y_n$. Moreover, in such case we may distinguish two possible situations: $\widehat{M}[y_{n,\lfloor \frac{n}{2} \rfloor - 1}] = \widehat{M}[x_n]$, or $\widehat{M}[y_{n,\lfloor \frac{n}{2} \rfloor - 1}] \neq \widehat{M}[x_n]$, but in last case $\widehat{M}[y_{n,\lfloor \frac{n}{2} \rfloor - 1}]$ "moves closer" to x_{ζ_1}, i.e. $D_1(x_{\zeta_1}, \widehat{M}[y_{n,\lfloor \frac{n}{2} \rfloor - 1}]) < D_1(x_{\zeta_1}, \widehat{M}[x_n])$.

Hence $\lfloor \frac{n}{2} \rfloor$ observations in $y_{n,\lfloor \frac{n}{2} \rfloor - 1}$ closest to $\widehat{M}[y_{n,\lfloor \frac{n}{2} \rfloor - 1}]$ have the distance ζ_1 from $\widehat{M}[x_n]$. The remaining elements in $y_{n,\lfloor \frac{n}{2} \rfloor - 1}$ are more distant from $\widetilde{Me}_H[y_{n,\lfloor \frac{n}{2} \rfloor - 1}]$, so we obtain exactly $\lfloor \frac{n}{2} \rfloor$ identical observations in $y_{n,\frac{1}{2}}$ and finally

$$\widetilde{IQR}_{D_1}(y_{n,\lfloor \frac{n}{2} \rfloor - 1}) = \max\{D_1(y_i, y_j) : y_i, y_j \in y_{n,\frac{1}{2}}\} = 0.$$

Thus, $\inf_{z_{n,\lfloor \frac{n}{2} \rfloor - 1}} \widetilde{IQR}_{D_1}(z_{n,\lfloor \frac{n}{2} \rfloor - 1}) = 0$ with any sample $z_{n,\lfloor \frac{n}{2} \rfloor - 1}$, so $\mathrm{fsbp}^-(\widetilde{IQR}_{D_1}, x_n) \leqslant \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}$.

Step 2: Secondly, we show that $\mathrm{fsbp}^-(\widetilde{IQR}_{D_1}, x_n) \geqslant \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}$.

Let $y_{n,k}$ be a sample obtained from x_n by substituting $k < \lfloor \frac{n}{2} \rfloor - 1$ observations with x_{ζ_1}. Under such construction of $y_{n,k}$ its median $\widehat{M}[y_{n,\lfloor \frac{n}{2} \rfloor - 1}]$ behaves as in the previous step of the proof (because all modified observation belong to $y_{n,\frac{1}{2}}$). Since $k < \lfloor \frac{n}{2} \rfloor - 1$ and $|y_{n,\frac{1}{2}}| = \lfloor \frac{n}{2} \rfloor$, there exist at least two observations $x^*, x^{**} \in x_n$ such that $x^*, x^{**} \in y_{n,\frac{1}{2}}$ (i.e. such observations that although remained unchanged but belong to $y_{n,\frac{1}{2}}$).

Let $\delta = D_1(x^*, x^{**}) > 0$ since we have assumed no identical observations in x_n. Hence

$$\delta = D_1(x^*, x^{**}) \leqslant \max\{D_1(y_i, y_j) : y_i, y_j \in y_{n,\frac{1}{2}}\} = \widetilde{IQR}_{D_1}(y_{n,k}).$$

Thus $\inf_{z_{n,k}} \widetilde{IQR}_{D_1}(z_{n,k}) \geqslant \delta > 0$ for $k < \lfloor \frac{n}{2} \rfloor - 1$ and any sample $z_{n,k}$, so $\mathrm{fsbp}^-(\widetilde{IQR}_{D_1}, x_n) \geqslant \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}$.

Summing up both Step 1 and Step 2 we obtain $\mathrm{fsbp}^-(\widetilde{IQR}_{D_1}, x_n) = \frac{\lfloor \frac{n}{2} \rfloor - 1}{n}$.

Step 3: Now, we will show that $\mathrm{fsbp}^+(\widetilde{IQR}_{D_1}, x_n) \leqslant \frac{\lceil \frac{n}{2} \rceil}{n}$.

Let

$$x_{max} = \{x \in x_{n,\frac{1}{2}} : \mathrm{wabl}(x) = \max_{x_i \in x_{n,\frac{1}{2}}} \mathrm{wabl}(x_i)\}$$

denote such observation which has "in the average" the greatest central point among observations in $x_{n,\frac{1}{2}}$. Moreover, let us fix $L \in \mathbb{R}$ such that $L > 0$.

We construct a sample $y_{n,\lceil \frac{n}{2} \rceil}$ by replacing $\lceil \frac{n}{2} \rceil$ observations in x_n as follows: observations belonging to $x_{n,\frac{1}{2}}$ remain without any change (i.e. $y_1 = x_{\xi_1}, ..., y_{\lfloor \frac{n}{2} \rfloor} = x_{\xi_{\lfloor \frac{n}{2} \rfloor}}$; one of those elements is x_{max}, further on denoted as y_{max}), and we modify observations from outside $x_{n,\frac{1}{2}}$ as follows: $\mathrm{wabl}\left(y_{\lfloor \frac{n}{2} \rfloor + 1}\right) = \mathrm{wabl}\,(y_{max}) + L$, $\mathrm{wabl}\left(y_{\lfloor \frac{n}{2} \rfloor + 2}\right) = \mathrm{wabl}\,(y_{max}) + 2L, \ldots,$ $\mathrm{wabl}\,(y_n) = \mathrm{wabl}\,(y_{max}) + \lceil \frac{n}{2} \rceil L$, however $\mathrm{ldev}_{y_k}(\alpha) = \mathrm{ldev}_{y_{max}}(\alpha)$ and $\mathrm{rdev}_{y_k}(\alpha) = \mathrm{rdev}_{y_{max}}(\alpha)$ for $k > \lfloor \frac{n}{2} \rfloor, \alpha \in [0, 1]$.

Now, for any $\forall \alpha \in [0, 1]$ and for any sample size n it is clear that $\mathrm{Me}(\mathrm{ldev}_{y_1}(\alpha), \ldots, \mathrm{ldev}_{y_n}(\alpha)) = \mathrm{ldev}_{y_{max}}(\alpha)$ or $\mathrm{Me}(\mathrm{rdev}_{y_1}(\alpha), \ldots, \mathrm{rdev}_{y_n}(\alpha)) = \mathrm{rdev}_{y_{max}}(\alpha)$, because there are $\lceil \frac{n}{2} \rceil + 1$ observations equal to y_{max} in $y_{n,\lceil \frac{n}{2} \rceil}$. On the other hand, if L is large enough then

$$\mathrm{Me}(\mathrm{wabl}(y_1), \ldots, \mathrm{wabl}(y_n)) = \mathrm{Me}(\ldots, \mathrm{wabl}(y_{max}), \mathrm{wabl}(y_{max}) + L, \ldots) = \mathrm{wabl}(y_{max}) + \tfrac{1}{2}L,$$

if n is even and $\mathrm{Me}(\mathrm{wabl}(y_1), \ldots, \mathrm{wabl}(y_n)) = \mathrm{wabl}(y_{max}) + L$, if n is odd. The aforementioned equalities hold because there are $\lfloor \frac{n}{2} \rfloor$ observations among observations in $y_{n,\lceil \frac{n}{2} \rceil}$ having "in the average" small central point and obtained without adding L. Moreover, $\mathrm{wabl}(y_{max})$ is the biggest central point in this group. On the other hand, the smallest central point in the subset of values obtained with addition L is $\mathrm{wabl}(y_{max}) + L$. Hence, for even sample size the median of the central points is obtained by the average of those two values mentioned above, while for odd sample size the median equals $\mathrm{wabl}(y_{max}) + L$.

Thus, for $\alpha \in [0, 1]$ we obtain

$$\left(\widehat{\mathrm{M}}(y_{n,\lceil \frac{n}{2} \rceil})\right)_\alpha = [\mathrm{wabl}(y_{max}) + \tfrac{1}{2}L - \mathrm{ldev}_{y_{max}}(\alpha), \mathrm{wabl}(y_{max}) + \tfrac{1}{2}L + \mathrm{rdev}_{y_{max}}(\alpha)],$$

if n is even and

$$\big(\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big)_\alpha = [\mathrm{wabl}(y_{max}) + L - \mathrm{ldev}_{y_{max}}(\alpha), \mathrm{wabl}(y_{max}) + L + \mathrm{rdev}_{y_{max}}(\alpha)].$$

if n is odd.

Let $C = \frac{1}{2}\int_{[0,1]} \big[\mathrm{rdev}_{y_{max}}(\alpha) - \mathrm{ldev}_{y_{max}}(\alpha)\big]d\alpha > 0$. By lemma 3.1 we conclude that

$$\mathrm{wabl}(\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})) = \mathrm{wabl}(y_{max}) + \frac{1}{2}L + C$$

if n is even and $\mathrm{wabl}(\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})) = \mathrm{wabl}\,(y_{max}) + L + C$ if n is odd. Moreover, $\mathrm{ldev}_{\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})}(\alpha) = \mathrm{ldev}_{y_{max}}(\alpha) + C$ and $\mathrm{rdev}_{\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})}(\alpha) = \mathrm{rdev}_{y_{max}}(\alpha) - C$, no matter whether n is even or odd. Consequently, if L is large enough then the significant component of the distance between y_i and $\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})$ is just the distance between their center points.

One may notice that observation $y_{\lfloor\frac{n}{2}\rfloor+1}$ belongs to $y_{n,\frac{1}{2}}$ because

$$\begin{aligned} D_1\big(y_{\lfloor\frac{n}{2}\rfloor+1}, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big) &= \big|\mathrm{wabl}(y_{max}) + L - (\mathrm{wabl}(y_{max}) + \tfrac{1}{2}L + C)\big| \\ &+ \frac{1}{2}\int_{[0,1]} \Big(\big|\mathrm{ldev}_{y_{max}}(\alpha) - (\mathrm{ldev}_{y_{max}}(\alpha) + C)\big| + \big|\mathrm{rdev}_{y_{max}}(\alpha) - (\mathrm{rdev}_{y_{max}}(\alpha) - C)\big|\Big)d\alpha \\ &= \big|\tfrac{1}{2}L - C\big| + \tfrac{1}{2}\int_{[0,1]} 2|C|d\alpha = \big|\tfrac{1}{2}L - C\big| + |C| = \tfrac{1}{2}L, \end{aligned}$$

if n is even and $D_1\big(y_{\lfloor\frac{n}{2}\rfloor+1}, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big) = 2C$, if n is odd. Indeed, if a fuzzy number is symmetrical, i.e. $\mathrm{rdev}_{y_{max}}(\alpha) = \mathrm{ldev}_{y_{max}}(\alpha)\ \forall\alpha \in [0,1]$, then $C = 0$, so besides $y_{\lfloor\frac{n}{2}\rfloor+1}$ observation $y_{n,\frac{1}{2}}$ will also belong to $y_{n,\frac{1}{2}}$.

Otherwise, if $\exists\alpha \in [0,1]$ such that $\mathrm{rdev}_{y_{max}}(\alpha) \neq \mathrm{ldev}_{y_{max}}(\alpha)$, then $D_1\big(y_{\lfloor\frac{n}{2}\rfloor+1}, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big) = \min\limits_{y_i \in y_{n,\lceil\frac{n}{2}\rceil}} \big\{D_1(y_i, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil}))\big\}$, because for even n and $y_a \in \{y_{\lfloor\frac{n}{2}\rfloor+2}, \dots y_n\}$ we have

$$\begin{aligned} D_1\big(y_a, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big) &= \big|\mathrm{wabl}(y_a) + aL - (\mathrm{wabl}(y_{max}) + \tfrac{1}{2}L + C)\big| \\ &+ \frac{1}{2}\int_{[0,1]} \Big(\big|\mathrm{ldev}_{y_a}(\alpha) - (\mathrm{ldev}_{y_{max}}(\alpha) + C)\big| + \big|\mathrm{rdev}_{y_a}(\alpha) - (\mathrm{rdev}_{y_{max}}(\alpha) - C)\big|\Big)d\alpha \\ &\geqslant |L(a - \tfrac{1}{2}) - C| \geq \tfrac{3}{2}L - C, \end{aligned}$$

for even n and $y_a \in \{y_1, \dots, y_{\lfloor\frac{n}{2}\rfloor}\}$ we obtain $D_1\big(y_a, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big) \geqslant \frac{1}{2}L + C$. On the other hand, if n is odd, we obtain $D_1\big(y_{\lfloor\frac{n}{2}\rfloor+1}, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big) = \min\limits_{y_i \in y_{n,\lceil\frac{n}{2}\rceil}} \big\{D_1\big(y_i, \widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})\big)\big\}$.

Hence, besides $y_{\lfloor\frac{n}{2}\rfloor+1}$ we find in $y_{n,\frac{1}{2}}$ only observations from $\big\{y_1, \dots, y_{\lfloor\frac{n}{2}\rfloor}\big\}$, because for L large enough their distance from $\widehat{\mathrm{M}}(y_{n,\lceil\frac{n}{2}\rceil})$ is smaller than for any member of $\big\{y_{\lfloor\frac{n}{2}\rfloor+2}, \dots y_n\big\}$. Let us take $y_b \in \big\{y_1, \dots, y_{\lfloor\frac{n}{2}\rfloor}\big\}$ such that $y_b \in y_{n,\frac{1}{2}}$ and let us denote $y_{\lfloor\frac{n}{2}\rfloor+1}$ and y_b as y^1 and y^2, respectively. Then

$$\begin{aligned} D_1(y^1, y^2) &= \big|\mathrm{wabl}(y_{max}) + L - \mathrm{wabl}(y_b)\big| \\ &+ \frac{1}{2}\int\limits_{[0,1]} \Big(\big|\mathrm{ldev}_{y_{max}}(\alpha) - \mathrm{ldev}_{y_b}(\alpha)\big| + \big|\mathrm{rdev}_{y_{max}}(\alpha) - \mathrm{rdev}_{y_b}(\alpha)\big|\Big)d\alpha \geqslant L, \end{aligned}$$

so we obtain

$$\widetilde{IQR}_{D_1}(y_{n,\lceil\frac{n}{2}\rceil}) = \max\big\{D_1(y_i, y_j) : y_i, y_j \in y_{n,\frac{1}{2}}\big\} \geqslant D_1(y^1, y^2) \geqslant L.$$

Leting $L \to \infty$ we obtain $\sup_{\widetilde{z}_{n,\lceil \frac{n}{2} \rceil}} \widetilde{IQR}_{D_1}(\widetilde{z}_{n,\lceil \frac{n}{2} \rceil}) = \infty$ for any sample $\widetilde{z}_{n,\lceil \frac{n}{2} \rceil}$ originated from x_n by replacing $\lceil \frac{n}{2} \rceil$ observations. Therefore $\text{fsbp}^+(\widetilde{IQR}_{D_1}, x_n) \leqslant \frac{\lceil \frac{n}{2} \rceil}{n}$.

Step 4: Finally, we will show that $\text{fsbp}^+(\widetilde{IQR}_{D_1}, x_n) \geqslant \frac{\lceil \frac{n}{2} \rceil}{n}$.

Let $y_{n,k}$ denote a sample obtained from x_n by modifying $k < \lceil \frac{n}{2} \rceil$ observations. Thus at least $\lfloor \frac{n}{2} \rfloor + 1$ original observations remain. Thus, even if those k observations become very large, such modification will not disturb $\widehat{\text{M}}(y_{n,k})$. It means that for a modification producing large enough observations their distance from $\widehat{\text{M}}(y_{n,k})$ be greater than the distance between remaining observations and $\widehat{\text{M}}(y_{n,k})$, which implies that those new observations will not appear in the set $y_{n,\frac{1}{2}}$ and, consequently, they have no impact on $\widetilde{IQR}_{D_1}$.

Suppose, $W = \max\{D_1(x_i, x_j) : x_i, x_j \in x_n\} < \infty$. Then for the modification producing large enough observations we have

$$W = \max\{d_H(x_i, x_j) : x_i, x_j \in x_n\} \geqslant \max\{D_1(y_i, y_j) : y_i, y_j \in y_{n,\frac{1}{2}}\} = \widetilde{IQR}_{D_1}(y_{n,k}).$$

Hence, for $k < \lceil \frac{n}{2} \rceil$ and for any sample $z_{n,k}$ we obtain $\sup_{z_{n,k}} \widetilde{IQR}_{D_1}(z_{n,k}) \leqslant W < \infty$, so $\text{fsbp}^+(\widetilde{IQR}_{D_1}, x_n) \geqslant \frac{\lceil \frac{n}{2} \rceil}{n}$.

Now, summing up both Step 3 and Step 4 we obtain $\text{fsbp}^+(\widetilde{IQR}_{D_1}, x_n) = \frac{\lceil \frac{n}{2} \rceil}{n}$, which completes the proof.

□

6. Conclusion

In this contribution we have considered dispersion measures for fuzzy random variables. In particular, we have suggested a generalization of the range and interquartile range for fuzzy samples. Moreover, we have determined the finite sample breakdown point for each of the introduced dispersion measures.

Obviously, many questions related to the dispersion measures for fuzzy data remain open. Firstly, one may ask which median works better as a reference measure of central tendency in the interquartile range. Moreover, the breakdown point for the interquartile range based on other distances than D_1 should be examined. Finally, the suggested dispersion measures based on multidistances are worth further study as well as their relation to the notion of statistical depth.

References

[1] Aumann, R.J., (1965), Integrals of set-valued functions, J. Math. Anal. Appl., 12, pp.1-12.

[2] Ban, A.I., Coroianu, L., Grzegorzewski, P., (2015), Fuzzy Numbers: Approximations, Ranking and Applications, Polish Academy of Sciences, Warsaw.

[3] Blanco-Fernández, A., Corral, N., González-Rodríguez, G., (2011), Estimation of a flexible simple linear model for interval data based on set arithmetic, Comput. Stat. Data Anal., 55, pp.2568-2578.

[4] Blanco-Fernández, A., Casals, M.R., Colubi, A., Corral, N., Garcia-Barzana, M., Gil, M.A., Gonzalez-Rodriguez, G., Lopez, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., de la Rosa de Saa S., Sinova, B., (2014), A distance-based statistic analysis of fuzzy number-valued data(with Rejoinder), Int. J. Approx. Reason., 55, pp.1487-1501, pp.1601-1605.

[5] Couso, I., Dubois, D., (2014), Statistical reasoning with set-valued information: Ontic vs. epistemic views, Int. J. Approx. Reason., 55, pp.1502–1518.

[6] de la Rosa de Saa S., Lubiano, M.A., Sinova, B., Filzmoser, P.F., (2017), Robust scale estimators for fuzzy data, Adv. Data Anal. Classif., 11, pp.731-758.

[7] Diamond, P., Kloeden, P., (1990), Metric spaces of fuzzy sets, Fuzzy Sets Syst., 35, pp.241-249.

[8] Donoho, D.L., (1982), Breakdown properties of multivariate location estimators, Pd.D. quantifying paper, Department of Statistics, Harvard University.

[9] Donoho, D.L., Huber, P.J., (1983), The notion of breakdown point, In: Bickel PJ, Doksum K, Hodges JL Jr (eds) A Festschrift for Erich L. Lehmann, Wadsworth, Belmont, pp.157-184.
[10] Dubois, D., Prade, H., (1978), Operations on fuzzy numbers, Int. J. Syst. Sci., 9, pp.613-626.
[11] Frachet, M., (1948), Les elements aleatoires de nature quelconque dan un espace distancie, Annales de Institut Henri Poincara, 10, pp.215-310.
[12] Gil, M. A., Lubiano, M. A., Montenegro, M., Lopez, M. T., (2002), Least squares fitting of an affine function and strength of association for interval-valued data, Metrika 56, pp.97-111.
[13] Gonzalez-Rodriguez, G., Colubi, A., Gil, M.A., (2012), Fuzzy data treated as functional data. A one-way ANOVA test approach, Comput. Stat. Data Anal., 56, pp.943-955.
[14] Grzegorzewski, P., (1998), Metrics and orders in space of fuzzy numbers, Fuzzy Sets Syst., 97, pp.83-94.
[15] Grzegorzewski, P., (1998), Statistical inference about the median from vague data, Control Cybern., 27, 447-464.
[16] Grzegorzewski, P., (2019), Measures of dispersion for interval data, In: Destercke S. et al. (Eds.), Uncertainty Modelling in Data Science, pp.91-98, Springer.
[17] Hampel, F.R., (1968), Contributions to the theory of robust estimation, Ph.D. Thesis, University of California, Berkeley.
[18] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., (1986), Robust Statistics: The Approach Based on Influence Functions, Wiley, 502p.
[19] Huber, P.J., (1981), Robust Statistics, Wiley, 370p.
[20] Hyndman, R.J., Fan, Y., (1996), Sample quantiles in statistical packages, Amer. Stat., 50, pp.361-365.
[21] Koacz, A., Grzegorzewski, P., (2016), Measures of dispersion for multidimensional data, Eur. J. Oper. Res., 251, pp.930-937.
[22] Kruse, R., Meyer, K.D., (1987), Statistics with Vague Data, D. Riedel Publishing Company, 279p.
[23] Lubiano, M.A., Gil, M.A., Lopez-Diaz, M., Lopez-Garca, M.T., (2000), The λ-mean squared dispersion associated with a fuzzy random variable, Fuzzy Sets Syst., 111, pp.307-317.
[24] Martin, J., Mayor, G., (2009), How separated Palma, Inca and Manacor are?, In: Proc. of the AGOP 2009, pp.195-200.
[25] Puri, M.L., Ralescu, D.A., (1985), The concept of normality for fuzzy random variables, Ann. Probab., 11, pp.1373-1379.
[26] Ramos-Guajardo, A.B., Lubiano, M.A., (2012), K-sample tests for equality of variances of random fuzzy sets, Comput. Stat. Data Anal., 56(4), pp.956-966.
[27] Sinova, B., (2018), Scale equivariant alternative for fuzzy M-estimators of location, In: E. Gil et al. (eds.), The Mathematics of the Uncertain, Springer, pp.733-743.
[28] Sinova, B., de la Rosa de Saa S., Gil, M.A., (2013), A generalized L1-type metric between fuzzy numbers for an approach to central tendency of fuzzy data, Inf. Sci., 242, pp.22-34.
[29] Sinova, B., de la Rosa de Saa S., Lubiano, M.A., Gil, M.A., (2015), An overview on the statistical central tendency for f uzzy data sets, Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 23, pp.105-132.
[30] Sinova, B., Gil, M.A., Colubi, A., Van, Aelst S., (2012), The median of a random fuzzy number. The 1-norm distance approach, Fuzzy Sets Syst., 200, pp.99-115.
[31] Sinova, B., Gil, M.A., Van, Aelst S., (2016), M-estimates of location for the robust central tendency of fuzzy data, IEEE Trans. Fuzzy Syst., 24(4), pp.945-956.
[32] Sinova, B., Van Aelst, S., (2018), Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function, Int. J. Approx. Reason., 93, pp.219-237.
[33] Sinova, B., Van Aelst, S., Teran, P., (2020), M-estimators and trimmed means: from Hilbert-valued to fuzzy set-valued data, Adv Data Anal. Classif., (to appear).
[34] Trutschnig, W., Gonzalez-Rodríguez, G., Colubi, A., Gil, M.A., (2009), A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread, Inf. Sci., 179, pp.3964-3972.
[35] Zadeh, L.A., (1965), Fuzzy sets, Inf. Control., 8, pp.338-353.
[36] Zielinski, R., (1977), Robustness: a quantitative approach, Bulletin de l'Academie Polonaise des Sciences. Sarie des Sciences Math., Astr. et Phys., 25(12), pp.1281-1286.

Przemyslaw Grzegorzewski - received his M.Sc. in mathematics from the University of Warsaw. Then he got his Ph.D. with distinction and D.Sc. (Habilitation) in computer science from the Systems Research Institute of the Polish Academy of Sciences. In 2018 he received the state title of Professor. He is currently a Full Professor at the Faculty of Mathematics and Information Science of Warsaw University of Technology and Systems Research Institute of the Polish Academy of Sciences. His research interests include mathematical statistics, statistical decisions with imprecise data, data mining, fuzzy sets, fuzzy logic, soft computing, statistical quality control, etc.

Katarzyna Gładek - received her master's degree from Mathematics and Information Science Faculty of Warsaw University of Technology in 2020. Her research interests include mathematical statistics and machine learning.