

## Robust scale estimators for fuzzy data

Sara de la Rosa de Saa · Marıa Asuncion  
Lubiano · Beatriz Sinova · Peter Filzmoser

Received: date / Accepted: date

**Abstract** Observations distant from the majority or deviating from the general pattern often appear in datasets. Classical estimates such as the sample mean or the sample variance can be substantially affected by these observations (outliers). Even a single outlier can have huge distorting influence. However, when one deals with real-valued data there exist robust measures/estimates of location and scale (dispersion) which reduce the influence of these atypical values and provide approximately the same results as the classical estimates applied to the typical data without outliers.

In real-life, data to be analyzed and interpreted are not always precisely defined and they cannot be properly expressed by using a numerical scale of measurement. Frequently, some of these imprecise data could be suitably described and modelled by considering a fuzzy rating scale of measurement.

In this paper, several well-known scale (dispersion) estimators in the real-valued case are extended for random fuzzy numbers (i.e., random mechanisms generating fuzzy-valued data), and some of their properties as estimators for dispersion are examined. Furthermore, their robust behaviour is analyzed using two powerful tools, namely, the finite sample breakdown point and the

---

Authors are very grateful for the insight comments from the reviewers of the original. The research in this paper has been partially supported by/benefited from Principality of Asturias Grants GRUPIN14-101, Research Contract E-33-2015-0040746 (this one for Sinova) and Severo Ochoa BP12012 (this one for De la Rosa de Saa), and the Spanish Ministry of Economy and Competitiveness Grant MTM2013-44212-P. Their financial support is gratefully acknowledged.

---

Sara de la Rosa de Saa · Marıa Asuncion Lubiano · Beatriz Sinova  
Departamento de Estadıstica e I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain  
E-mail: rosasara@uniovi.es, lubiano@uniovi.es, sinovabeatriz@outlook.es

Sara de la Rosa de Saa · Peter Filzmoser  
Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, 1040 Vienna, Austria  
E-mail: P.Filzmoser@tuwien.ac.at

sensitivity curves. Simulations, including empirical bias curves, are performed to complete the study.

**Keywords** finite sample breakdown point · empirical bias curves · fuzzy numbers · random fuzzy numbers · robustness · scale estimation · sensitivity curves

## 1 Introduction

Imprecise data can be found in many real-life situations. Fields like Social and Biomedical Sciences or Engineering have often to deal with this kind of data. For instance, the valuation/rating of the employee productivity, the customer satisfaction, the water quality or the technological impact, are matters for which values correspond to intrinsically imprecise data. This type of data can be often properly expressed and modelled by means of fuzzy numbers.

The concept of the so-called random fuzzy numbers (RFNs for short, see Puri and Ralescu [23] for the seminal reference about), has been introduced to formalize the random mechanism generating fuzzy number-valued data within a probabilistic setting.

In summarizing or representing the central tendency or location of an RFN, some measures have been suggested, namely, the so-called Aumann-type mean (see Puri and Ralescu [23]) and two recently introduced extensions of the median based on  $L^1$ -type metrics between fuzzy numbers (more concretely, the 1-norm and the wabl/ldev/rdev ones, see Sinova *et al.* [26,27]).

The measurement of the variability/dispersion of an RFN is also a useful complementary tool to summarize its distribution. In a fuzzy setting the variability of the values of an RFN has been measured by means of the Fréchet variance, defined in terms of a convenient  $L^2$ -type metric (see, for instance, Lubiano *et al.* [18], Ramos-Guajardo and Lubiano [24], Blanco-Fernández *et al.* [1,2]).

An important and valuable feature of an (either location or scale) estimator is to be robust. The lack of robustness of an estimator makes it to be sensitive to the influence of extreme/atypical values (outliers) among data. Sinova *et al.* [26,27] have shown that, as for the real-valued case, whereas the Aumann-type mean is highly affected by the presence of outliers, the 1-norm and the wabl/ldev/rdev medians are two robust alternatives in estimating location.

The aim of this contribution is to introduce some scale measures for RFNs on the basis of two  $L^1$ -type metrics, and to analyze their robustness, that is, their sensitivity to either changes of values or the presence of outliers and inliers (usually repeated values in the data context) among data, as well as to compare them with the standard deviation measure. The finite sample breakdown point, the empirical bias curves and the sensitivity curves will be the employed tools to carry out this study.

In Section 2 the preliminaries on fuzzy numbers, arithmetic and metrics between them, and the concept of RFN and related location measures are recalled. Section 3 extends some scale measures for fuzzy data, analyzing some

of their relevant properties as dispersion measures. In Section 4 we focus on the study of the robustness of the measures through the finite sample breakdown point of the scale estimators considered in the previous section; simulations are also performed to represent the corresponding empirical bias curves. Section 5 presents the sensitivity curves associated with the scale estimators. Finally, some future research directions are commented.

## 2 Preliminaries about fuzzy data associated with random experiments

As we have already commented, in dealing with imprecise data associated with random experiments, fuzzy numbers become often an appropriate concept to express and model these data (see, for instance, De la Rosa de Sáa *et al.* [6], Blanco-Fernández *et al.* [1, 2]). Actually, fuzzy numbers determine a continuous scale in which each element has an intuitive meaning, although elements cannot be one-to-one translated into words of a natural language that should be necessarily finite. Furthermore, this scale is very rich and expressive and it can be well-handled mathematically through a suitable arithmetic. Although this arithmetic does not inherit all the properties of the arithmetic for real numbers, the use of different metrics helps often to overcome the associated drawbacks. The preliminary concepts about are to be recalled in Subsection 2.1.

To analyze statistically fuzzy data the preceding concepts should be accompanied with the modelling of the random mechanisms generating them. A well-supported model for this purpose is the one given within the probabilistic setting by random fuzzy numbers. Subsection 2.2 recalls this concept along with some of the most relevant known summary location measures associated with its induced distribution.

### 2.1 Fuzzy data, arithmetic and metrics

Fuzzy numbers (also referred to by some authors as fuzzy intervals) are formalized as follows:

**Definition 1** A (bounded) **fuzzy number** is a mapping  $\tilde{U} : \mathbb{R} \rightarrow [0, 1]$  such that for all  $\alpha \in [0, 1]$  the  $\alpha$ -level set defined as

$$\tilde{U}_\alpha = \begin{cases} \{x \in \mathbb{R} : \tilde{U}(x) \geq \alpha\} & \text{if } \alpha \in (0, 1] \\ \text{cl}\{x \in \mathbb{R} : \tilde{U}(x) > 0\} & \text{if } \alpha = 0 \end{cases}$$

with ‘cl’ denoting the closure of the set, is a nonempty compact interval. For each  $x \in \mathbb{R}$ , the value  $\tilde{U}(x)$  can be interpreted as the ‘degree of compatibility of  $x$  with the property defined by  $\tilde{U}$ ’. The space of (bounded) fuzzy numbers will be denoted by  $\mathcal{F}_c^*(\mathbb{R})$ .

Real numbers can be viewed as special fuzzy numbers, so that each real number  $x$  can be identified with the indicator function of the corresponding singleton  $\mathbb{1}_{\{x\}}$ .

When fuzzy numbers are considered to model experimental data, statistics to analyze them are frequently based on two arithmetical operations, namely, the sum and the product by scalars. The common way to extend the sum and the product by a scalar from  $\mathbb{R}$  to  $\mathcal{F}_c^*(\mathbb{R})$  is to use Zadeh's extension principle [30], which is equivalent to consider level-wise the usual interval arithmetic. More concretely,

**Definition 2** Given  $\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ , the **sum** of  $\tilde{U}$  and  $\tilde{V}$  is the fuzzy number  $\tilde{U} + \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$  such that for each  $\alpha \in [0, 1]$

$$(\tilde{U} + \tilde{V})_\alpha = \text{Minkowski sum of } \tilde{U}_\alpha \text{ and } \tilde{V}_\alpha = [\inf \tilde{U}_\alpha + \inf \tilde{V}_\alpha, \sup \tilde{U}_\alpha + \sup \tilde{V}_\alpha].$$

**Definition 3** Given  $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$  and  $\gamma \in \mathbb{R}$ , the **product** of  $\tilde{U}$  by the scalar  $\gamma$  is the fuzzy number  $\gamma \cdot \tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$  such that for each  $\alpha \in [0, 1]$

$$(\gamma \cdot \tilde{U})_\alpha = \begin{cases} [\gamma \cdot \inf \tilde{U}_\alpha, \gamma \cdot \sup \tilde{U}_\alpha] & \text{if } \gamma \geq 0 \\ [\gamma \cdot \sup \tilde{U}_\alpha, \gamma \cdot \inf \tilde{U}_\alpha] & \text{otherwise.} \end{cases}$$

A remarkable fact and distinguishing feature in contrast to the real-valued case, is that when  $\mathcal{F}_c^*(\mathbb{R})$  is endowed with the preceding arithmetic it does not have a linear but a conical (i.e. semilinear) structure. Thus,  $\tilde{U} + (-1) \cdot \tilde{U} \neq \mathbb{1}_{\{0\}}$  = neutral element for the fuzzy sum, but in this case  $\tilde{U}$  reduces to a real number.

Due to such a nonlinearity, one cannot state a definition for the difference between fuzzy numbers that is always well-defined and simultaneously satisfies that  $(\tilde{U} - \tilde{V}) + \tilde{V} = \tilde{U}$  whatever  $\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$  may be. This crucial drawback can be substantially overcome in developing statistics with fuzzy data by involving suitable metrics. In this paper, we will make use of the following ones:

**Definition 4** (Diamond and Kloeden [7]) Given  $\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ ,

- the mapping  $\rho_1 : \mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R}) \rightarrow [0, +\infty)$  defined as

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

is called the **1-norm distance** between  $\tilde{U}$  and  $\tilde{V}$ ;

- the mapping  $\rho_2 : \mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R}) \rightarrow [0, +\infty)$  defined as

$$\rho_2(\tilde{U}, \tilde{V}) = \sqrt{\frac{1}{2} \int_{(0,1]} \left( \left[ \inf \tilde{U}_\alpha - \inf \tilde{V}_\alpha \right]^2 + \left[ \sup \tilde{U}_\alpha - \sup \tilde{V}_\alpha \right]^2 \right) d\alpha}$$

is called the **2-norm distance** between  $\tilde{U}$  and  $\tilde{V}$ .

**Definition 5** (Sinova *et al.* [27]) Given  $\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ , the mapping  $\mathcal{D}_1 : \mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R}) \rightarrow [0, +\infty)$  given by

$$\begin{aligned} \mathcal{D}_1(\tilde{U}, \tilde{V}) &= |\text{wabl}(\tilde{U}) - \text{wabl}(\tilde{V})| \\ &+ \frac{1}{2} \int_{[0,1]} (|\text{ldev}_{\tilde{U}}(\alpha) - \text{ldev}_{\tilde{V}}(\alpha)| + |\text{rdev}_{\tilde{U}}(\alpha) - \text{rdev}_{\tilde{V}}(\alpha)|) d\alpha, \end{aligned}$$

where

$$\text{wabl}(\tilde{U}) = \int_{[0,1]} \text{mid} \tilde{U}_\alpha d\alpha,$$

$$\text{ldev}_{\tilde{U}}(\alpha) = \text{wabl}(\tilde{U}) - \inf \tilde{U}_\alpha, \quad \text{rdev}_{\tilde{U}}(\alpha) = \sup \tilde{U}_\alpha - \text{wabl}(\tilde{U}),$$

with  $\text{mid} \tilde{U}_\alpha$  denoting the center/mid-point of  $\tilde{U}_\alpha$ , is called the **wabl/ldev/rdev-based  $L^1$  metric** between  $\tilde{U}$  and  $\tilde{V}$ .

Metrics  $\rho_1$  and  $\mathcal{D}_1$  are  $L^1$ -type ones on  $\mathcal{F}_c^*(\mathbb{R})$ , whereas  $\rho_2$  is  $L^2$ -type, and the three corresponding metric spaces are separable. Moreover, through the support function of fuzzy sets (see Puri and Ralescu [22]) and with the above recalled arithmetic the metric spaces  $(\mathcal{F}_c^*(\mathbb{R}), \rho_1)$  and  $(\mathcal{F}_c^*(\mathbb{R}), \mathcal{D}_1)$  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, and  $(\mathcal{F}_c^*(\mathbb{R}), \rho_2)$  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.

## 2.2 Random fuzzy numbers and some relevant central tendency measures

To formalize the random mechanisms generating fuzzy data within a probabilistic setting, the following notion is considered

**Definition 6** (Puri and Ralescu [23]) Given a probability space  $(\Omega, \mathcal{A}, P)$ , an associated **random fuzzy number** (for short RFN) is a mapping  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c^*(\mathbb{R})$  such that for all  $\alpha \in [0, 1]$  the real-valued mappings  $\inf \mathcal{X}_\alpha$  and  $\sup \mathcal{X}_\alpha$  are random variables, where  $\mathcal{X}_\alpha : \Omega \rightarrow \mathcal{P}(\mathbb{R})$  is such that  $\mathcal{X}_\alpha(\omega) = (\mathcal{X}(\omega))_\alpha$ .

*Remark 1* Equivalently,  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c^*(\mathbb{R})$  is said to be an RFN if  $\mathcal{X}$  is a Borel-measurable mapping w.r.t. the Borel  $\sigma$ -field generated on  $\mathcal{F}_c^*(\mathbb{R})$  by the topology induced by several different metrics, like those in Definitions 4 and 5. This Borel-measurability ensures that one can properly refer to the *distribution induced by an RFN*, the *stochastic independence of RFNs*, and so on, without needing to state expressly these notions.

Location measures are the first logical attempt when one wants to summarize the distribution of either a random fuzzy number (population measure) or a sample of observations from it (sample measure). The best known location measure is the so-called Aumann-type mean value for random fuzzy numbers, which extends the mean of a random variable and is formalized as follows:

**Definition 7** (Puri and Ralescu [23]) Let  $\mathcal{X}$  be a random fuzzy number associated with the probability space  $(\Omega, \mathcal{A}, P)$  which is integrably bounded, that is,  $E(\max\{|\inf \mathcal{X}_0|, |\sup \mathcal{X}_0|\}) < \infty$ . The **(population) Aumann-type mean** or **expected value** of  $\mathcal{X}$  is the fuzzy number  $\tilde{E}(\mathcal{X}) \in \mathcal{F}_c^*(\mathbb{R})$  such that for each  $\alpha \in [0, 1]$

$$(\tilde{E}(\mathcal{X}))_\alpha = [E(\inf \mathcal{X}_\alpha), E(\sup \mathcal{X}_\alpha)]$$

with  $E$  denoting the expected value of a real-valued random variable.

In particular, if  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  is a sample of observations from  $\mathcal{X}$ , the **(sample) Aumann-type mean** is the fuzzy number  $\tilde{\tilde{\mathbf{x}}}_n$  given for all  $\alpha \in [0, 1]$  by

$$(\tilde{\tilde{\mathbf{x}}}_n)_\alpha = \left( \frac{1}{n} \cdot (\tilde{x}_1 + \dots + \tilde{x}_n) \right)_\alpha = \left[ \frac{1}{n} \sum_{i=1}^n \inf(\tilde{x}_i)_\alpha, \frac{1}{n} \sum_{i=1}^n \sup(\tilde{x}_i)_\alpha \right].$$

Two recently introduced location measures extending the median of a random variable are the following:

**Definition 8** (Sinova *et al.* [26]) Let  $\mathcal{X}$  be a random fuzzy number. The **(population) 1-norm median** of  $\mathcal{X}$  is the fuzzy number  $\tilde{\text{Me}}(\mathcal{X})$  such that for each  $\alpha \in [0, 1]$

$$(\tilde{\text{Me}}(\mathcal{X}))_\alpha = [\text{Me}(\inf \mathcal{X}_\alpha), \text{Me}(\sup \mathcal{X}_\alpha)],$$

with  $\text{Me}$  denoting the median of a real-valued random variable and where, in case  $\text{Me}(\inf \mathcal{X}_\alpha)$  or  $\text{Me}(\sup \mathcal{X}_\alpha)$  are non-unique, we will follow the most usual convention, that is, we will consider the midpoint of the interval of medians.

In particular, if  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  is a sample of observations from  $\mathcal{X}$ , the **(sample) 1-norm median** is the fuzzy number  $\widehat{\tilde{\text{Me}}}(\tilde{\mathbf{x}}_n)$  given for all  $\alpha \in [0, 1]$  by

$$\left( \widehat{\tilde{\text{Me}}}(\tilde{\mathbf{x}}_n) \right)_\alpha = [\text{Me}\{\inf(\tilde{x}_1)_\alpha, \dots, \inf(\tilde{x}_n)_\alpha\}, \text{Me}\{\sup(\tilde{x}_1)_\alpha, \dots, \sup(\tilde{x}_n)_\alpha\}].$$

**Definition 9** (Sinova *et al.* [27]) Let  $\mathcal{X}$  be a random fuzzy number. The **(population) wabl/ldev/rdev median** of  $\mathcal{X}$  is the fuzzy number  $\tilde{\text{M}}(\mathcal{X})$  such that for each  $\alpha \in [0, 1]$

$$(\tilde{\text{M}}(\mathcal{X}))_\alpha = [\text{Me}(\text{wabl}(\mathcal{X})) - \text{Me}(\text{ldev}_\mathcal{X}(\alpha)), \text{Me}(\text{wabl}(\mathcal{X})) + \text{Me}(\text{rdev}_\mathcal{X}(\alpha))],$$

where in case  $\text{Me}(\text{wabl}(\mathcal{X}))$ ,  $\text{Me}(\text{ldev}_\mathcal{X}(\alpha))$  or  $\text{Me}(\text{rdev}_\mathcal{X}(\alpha))$  are non-unique the most usual convention is considered.

In particular, if  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  is a sample of observations from  $\mathcal{X}$ , the *(sample) wabl/ldev/rdev median* is the fuzzy number  $\tilde{M}(\tilde{\mathbf{x}}_n)$  given for all  $\alpha \in [0, 1]$  by

$$\left( \tilde{M}(\tilde{\mathbf{x}}_n) \right)_\alpha = \left[ \text{Me}\{\text{wabl}(\tilde{x}_1), \dots, \text{wabl}(\tilde{x}_n)\} - \text{Me}\{\text{ldev}_{\tilde{x}_1}(\alpha), \dots, \text{ldev}_{\tilde{x}_n}(\alpha)\}, \right. \\ \left. \text{Me}\{\text{wabl}(\tilde{x}_1), \dots, \text{wabl}(\tilde{x}_n)\} + \text{Me}\{\text{rdev}_{\tilde{x}_1}(\alpha), \dots, \text{rdev}_{\tilde{x}_n}(\alpha)\} \right].$$

It should be pointed out that the preceding location measures preserve several valuable properties from the real-valued case.

**Proposition 1** (Puri and Ralescu [23], Sinova *et al.* [26], Sinova *et al.* [27])  *$\tilde{E}$ ,  $\tilde{Me}$  and  $\tilde{M}$  are equivariant under affine transformations on  $\mathcal{F}_c^*(\mathbb{R})$ , that is, if  $\gamma \in \mathbb{R}$ ,  $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$  and  $\tilde{E}(\mathcal{X})$  exists, then*

$$\tilde{E}(\gamma \cdot \mathcal{X} + \tilde{U}) = \gamma \cdot \tilde{E}(\mathcal{X}) + \tilde{U},$$

$$\tilde{Me}(\gamma \cdot \mathcal{X} + \tilde{U}) = \gamma \cdot \tilde{Me}(\mathcal{X}) + \tilde{U}, \quad \tilde{M}(\gamma \cdot \mathcal{X} + \tilde{U}) = \gamma \cdot \tilde{M}(\mathcal{X}) + \tilde{U}.$$

**Proposition 2** (Körner [14], Lubiano *et al.* [18], Sinova *et al.* [26], Sinova *et al.* [27]) *Let  $\mathcal{X}$  be a random fuzzy number. Then,*

$$\tilde{E}(\mathcal{X}) = \arg \min_{\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})} E \left( \left[ \rho_2(\mathcal{X}, \tilde{U}) \right]^2 \right),$$

$$\tilde{Me}(\mathcal{X}) = \arg \min_{\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})} E \left( \rho_1(\mathcal{X}, \tilde{U}) \right), \quad \tilde{M}(\mathcal{X}) = \arg \min_{\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})} E \left( \mathcal{D}_1(\mathcal{X}, \tilde{U}) \right),$$

*whenever the involved expectations exist.*

### 3 Scale measures for random fuzzy numbers

In the preceding section some location measures for random fuzzy numbers have been recalled. In addition to central tendency, another useful summary tool associated with the distribution of a random element is the measurement of its dispersion/variability, the so-called scale measures being the ones quantifying it.

In this section, several scale measures are introduced for random fuzzy numbers, the measures being based on the metrics and the location measures recalled in Section 2. In formalizing these measures, Fréchet's approach [10] is to be considered. In accordance with this approach the scale measures can be interpreted as a measure of the 'mean error' in approximating/estimating the values of the RFN by means of a fuzzy number. The scale measures we are going to present are assumed to be given in the same 'units' of the RFN.

Let  $\mathcal{X}$  be an RFN,  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  be a sample of observations from  $\mathcal{X}$  and consider the metrics between fuzzy numbers  $\rho_2$  and  $D_1 \in \{\rho_1, \mathcal{D}_1\}$ .

**Definition 10** The **(population) standard deviation** is the real number  $\text{SD}(\mathcal{X})$ , if it exists, given by

$$\text{SD}(\mathcal{X}) = \sqrt{E \left( \left[ \rho_2(\mathcal{X}, \tilde{E}(\mathcal{X})) \right]^2 \right)}.$$

In particular, the **(sample) standard deviation** is the real number  $\widehat{\text{SD}}(\tilde{\mathbf{x}}_n)$  given by

$$\widehat{\text{SD}}(\tilde{\mathbf{x}}_n) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[ \rho_2(\tilde{x}_i, \tilde{\bar{\mathbf{x}}}_n) \right]^2}.$$

*Remark 2* It should be pointed out that the standard deviation of an RFN corresponds to the squared root of the variance introduced by K3rner [14] (see Lubiano *et al.* [18], K3rner and N3ther [15] and Gonz3lez-Rodr3guez *et al.* [12], for the definition based on generalized metrics and more general fuzzy set-valued random elements).

**Definition 11** The **(population) average distance deviation about the mean** is the real number  $D_1\text{-ADD}(\mathcal{X})$ , if it exists, given by

$$D_1\text{-ADD}(\mathcal{X}) = E \left( D_1(\mathcal{X}, \tilde{E}(\mathcal{X})) \right).$$

In particular, the **(sample) average distance deviation about the mean** is the real number  $\widehat{D_1\text{-ADD}}(\tilde{\mathbf{x}}_n)$  given by

$$\widehat{D_1\text{-ADD}}(\tilde{\mathbf{x}}_n) = \frac{1}{n} \sum_{i=1}^n D_1(\tilde{x}_i, \tilde{\bar{\mathbf{x}}}_n).$$

**Definition 12** The **(population) median distance deviation about the  $D_1$ -median** is the real number  $D_1\text{-MDD}(\mathcal{X})$ , if it exists, given by

$$D_1\text{-MDD}(\mathcal{X}) = \begin{cases} \text{Me} \left( \rho_1(\mathcal{X}, \tilde{\text{Me}}(\mathcal{X})) \right) & \text{if } D_1 = \rho_1 \\ \text{Me} \left( \mathcal{D}_1(\mathcal{X}, \tilde{\text{M}}(\mathcal{X})) \right) & \text{if } D_1 = \mathcal{D}_1 \end{cases}$$

In particular, the **(sample) median distance deviation about the  $D_1$ -median** is the real number  $\widehat{D_1\text{-MDD}}(\tilde{\mathbf{x}}_n)$  given by

$$\widehat{D_1\text{-MDD}}(\tilde{\mathbf{x}}_n) = \begin{cases} \text{Me} \left\{ \rho_1 \left( \tilde{x}_1, \widehat{\text{Me}}(\tilde{\mathbf{x}}_n) \right), \dots, \rho_1 \left( \tilde{x}_n, \widehat{\text{Me}}(\tilde{\mathbf{x}}_n) \right) \right\} & \text{if } D_1 = \rho_1 \\ \text{Me} \left\{ \mathcal{D}_1 \left( \tilde{x}_1, \widehat{\text{M}}(\tilde{\mathbf{x}}_n) \right), \dots, \mathcal{D}_1 \left( \tilde{x}_n, \widehat{\text{M}}(\tilde{\mathbf{x}}_n) \right) \right\} & \text{if } D_1 = \mathcal{D}_1 \end{cases}$$

where in case Me is not unique, the usual convention of the mid-point of possible medians will be applied.

Before analyzing some relevant properties of these measures which are inherited from the real-valued case, we are going to prove that they are well-defined whenever they exist, because the involved random distances are real-valued random variables, and they actually extend well-known scale measures for real-valued random variables. Thus,

**Proposition 3** *Let  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c^*(\mathbb{R})$  be an RFN associated with the probability space  $(\Omega, \mathcal{A}, P)$ , and let  $\tilde{U}$  be an arbitrary fuzzy number. Then, the mappings  $\rho_2(\mathcal{X}, \tilde{U}) : \Omega \rightarrow \mathbb{R}$ ,  $\rho_1(\mathcal{X}, \tilde{U}) : \Omega \rightarrow \mathbb{R}$  and  $\mathcal{D}_1(\mathcal{X}, \tilde{U}) : \Omega \rightarrow \mathbb{R}$  are  $(\mathcal{A}, \mathcal{B}_{\mathbb{R}})$ -measurable (i.e., they are real-valued random variables).*

*Proof* Let  $([0, 1], \mathcal{M}_{[0,1]}, \lambda)$  denote the Lebesgue measure space on  $[0, 1]$ . Mappings  $h_1, h_2 : \Omega \times [0, 1] \rightarrow \mathbb{R}$  with  $h_1(\omega, \alpha) = \inf \mathcal{X}_\alpha(\omega)$  and  $h_2(\omega, \alpha) = \sup \mathcal{X}_\alpha(\omega)$  are  $\mathcal{A} \otimes \mathcal{M}_{[0,1]}$ -measurable (see López-Díaz and Gil [16]). Consequently,  $h_3 : \Omega \times [0, 1] \rightarrow \mathbb{R}$  with  $h_3(\omega, \alpha) = \text{mid } \mathcal{X}_\alpha(\omega)$  is also  $\mathcal{A} \otimes \mathcal{M}_{[0,1]}$ -measurable. Since  $\lambda$  is a finite measure, on the basis of the Fubini Theorem one can guarantee that the function  $g : \Omega \times [0, 1] \rightarrow \mathbb{R}$  with  $g(\omega, \alpha) = \text{wabl}(\mathcal{X}(\omega))$  whatever  $\alpha \in [0, 1]$  may be is  $\mathcal{A} \otimes \mathcal{M}_{[0,1]}$ -measurable.

On the other hand, mappings  $h_1^*, h_2^* : \Omega \times [0, 1] \rightarrow \mathbb{R}$  with  $h_1^*(\omega, \alpha) = \inf \tilde{U}_\alpha$  and  $h_2^*(\omega, \alpha) = \sup \tilde{U}_\alpha$  are left-continuous with respect to  $\alpha$  on  $[0, 1]$  and do not depend on  $\omega$ , so that they are  $\mathcal{A} \otimes \mathcal{M}_{[0,1]}$ -measurable. Trivially, the constant function  $g^* : \Omega \times [0, 1] \rightarrow \mathbb{R}$  with  $g^*(\omega, \alpha) = \text{wabl}(\tilde{U})$  whatever  $\omega \in \Omega$  and  $\alpha \in [0, 1]$  maybe is  $\mathcal{A} \otimes \mathcal{M}_{[0,1]}$ -measurable.

As a consequence from these conclusions, the functions  $[h_1(\cdot, \cdot) - h_1^*(\cdot, \cdot)]^2 + [h_2(\cdot, \cdot) - h_2^*(\cdot, \cdot)]^2$ ,  $|h_1(\cdot, \cdot) - h_1^*(\cdot, \cdot)| + |h_2(\cdot, \cdot) - h_2^*(\cdot, \cdot)|$ ,  $|g(\cdot, \cdot) - g^*(\cdot, \cdot)|$  and  $|g(\cdot, \cdot) - h_1(\cdot, \cdot) - (g^*(\cdot, \cdot) - h_1^*(\cdot, \cdot))| + |h_2(\cdot, \cdot) - g(\cdot, \cdot) - (h_2^*(\cdot, \cdot) - g^*(\cdot, \cdot))|$  are  $\mathcal{A} \otimes \mathcal{M}_{[0,1]}$ -measurable. On the basis of the Fubini Theorem, and because of  $\lambda$  being a finite measure, we have that  $\rho_2(\mathcal{X}, \tilde{U})$ ,  $\rho_1(\mathcal{X}, \tilde{U})$  and  $\mathcal{D}_1(\mathcal{X}, \tilde{U})$  are  $(\mathcal{A}, \mathcal{B}_{\mathbb{R}})$ -measurable.  $\square$

On the basis of the preceding result, the scale measures introduced in this section are well-defined whenever they exist. Moreover, they extend well-known measures for real-valued random variables, as it is now to be demonstrated.

**Proposition 4** *Let  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c^*(\mathbb{R})$  be an RFN associated with the probability space  $(\Omega, \mathcal{A}, P)$ . If  $\mathcal{X} = \mathbb{1}_{\{X\}}$  a.s.  $[P]$  for a real-valued random variable  $X$  associated with  $(\Omega, \mathcal{A}, P)$ , then  $\text{SD}(\mathcal{X})$ ,  $D_1\text{-ADD}(\mathcal{X})$  and  $D_1\text{-MDD}(\mathcal{X})$  coincide with the corresponding measures for the real-valued case.*

*Proof* This is immediate taking into account that

- the Aumann-type mean, the 1-norm median and the wabl/ldev/rdev median coincide in such situation with the corresponding measures for the real-valued case
- and the  $\rho_2$  and  $D_1$  metrics are generalizations of the Euclidean distance in  $\mathbb{R}$ .  $\square$

The scale measures in Definitions 10-12 preserve many interesting properties for real-valued random variables. More concretely,

**Proposition 5** *SD,  $D_1$ -ADD and  $D_1$ -MDD are nonnegative operators.*

**Proposition 6** *SD( $\mathcal{X}$ ) = 0, or  $D_1$ -ADD( $\mathcal{X}$ ) = 0, holds if, and only if, the RFN  $\mathcal{X}$  is degenerate at a fuzzy number (i.e., there exists a fuzzy number  $\tilde{U}$  such that  $\mathcal{X} = \tilde{U}$  a.s. [P]). Furthermore, if the RFN  $\mathcal{X}$  is degenerate at a fuzzy number, then  $D_1$ -MDD( $\mathcal{X}$ ) = 0.*

*Proof* Indeed, because of the non-negativity of the distances between fuzzy numbers, one can conclude that SD( $\mathcal{X}$ ) = 0 or  $D_1$ -ADD( $\mathcal{X}$ ) = 0 if, and only if,  $D(\mathcal{X}, \tilde{E}(\mathcal{X})) = 0$  a.s. [P] whatever  $D \in \{\rho_2, \rho_1, \mathcal{D}_1\}$  may be, that is, if and only if  $\mathcal{X} = \tilde{E}(\mathcal{X})$  a.s. [P].

On the other hand, if  $\mathcal{X} = \tilde{U}$  a.s. [P], then  $D_1(\mathcal{X}, \tilde{U}) = 0$  a.s. [P], whence  $D_1$ -MDD( $\mathcal{X}$ ) = 0.  $\square$

*Remark 3* As for the real-valued case,  $D_1$ -MDD( $\mathcal{X}$ ) = 0 does not generally imply that the RFN  $\mathcal{X}$  is degenerate at a fuzzy number. For instance, if  $\mathcal{X}$  is a nondegenerate random fuzzy number taking on the triangular fuzzy values  $\text{Tri}(0, 1, 2)$  and  $\text{Tri}(1, 2, 3)$  with probabilities  $2/3$  and  $1/3$ , respectively (where,  $(\text{Tri}(a, b, c))_\alpha = [\alpha \cdot b + (1 - \alpha) \cdot a, \alpha \cdot b + (1 - \alpha) \cdot c]$ ), then  $\tilde{\text{Me}}(\mathcal{X}) = \tilde{\text{M}}(\mathcal{X}) = \text{Tri}(0, 1, 2)$ , whence  $D_1$ -MDD( $\mathcal{X}$ ) =  $\text{Me}\{0, 0, 1\} = 0$ .

**Proposition 7** *SD,  $D_1$ -ADD and  $D_1$ -MDD are shift invariant and scale equivariant. That is, if  $\gamma \in \mathbb{R}$ ,  $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$  and  $\mathcal{X}$  is an RFN for which the involved scale measures exist, then,*

$$\begin{aligned} \text{SD}(\gamma \cdot \mathcal{X} + \tilde{U}) &= |\gamma| \cdot \text{SD}(\mathcal{X}), & D_1\text{-ADD}(\gamma \cdot \mathcal{X} + \tilde{U}) &= |\gamma| \cdot D_1\text{-ADD}(\mathcal{X}), \\ D_1\text{-MDD}(\gamma \cdot \mathcal{X} + \tilde{U}) &= |\gamma| \cdot D_1\text{-MDD}(\mathcal{X}). \end{aligned}$$

*Proof* Indeed, whatever the metric  $D \in \{\rho_2, \rho_1, \mathcal{D}_1\}$  and the location measure  $\tilde{C}(\mathcal{X}) \in \{\tilde{E}(\mathcal{X}), \tilde{\text{Me}}(\mathcal{X}), \tilde{\text{M}}(\mathcal{X})\}$  may be, because of the properties of the metrics and the location measures we have that

$$D(\gamma \cdot \mathcal{X} + \tilde{U}, \tilde{C}(\gamma \cdot \mathcal{X} + \tilde{U})) = D(\gamma \cdot \mathcal{X} + \tilde{U}, \gamma \cdot \tilde{C}(\mathcal{X}) + \tilde{U}) = |\gamma| \cdot D(\mathcal{X}, \tilde{C}(\mathcal{X})),$$

whence the proof is immediately concluded.  $\square$

**Proposition 8** *Let  $\mathcal{X}$  be an RFN associated with the probability space  $(\Omega, \mathcal{A}, P)$ , and let  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple random sample from  $\mathcal{X}$ .*

*i) If  $\tilde{E}(\mathcal{X})$  and SD( $\mathcal{X}$ ) exist, then the statistic  $\widehat{\text{SD}}(\mathcal{X}_1, \dots, \mathcal{X}_n)$  is a strongly consistent estimator of SD( $\mathcal{X}$ ), that is,*

$$\lim_{n \rightarrow \infty} \widehat{\text{SD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = \text{SD}(\mathcal{X}) \quad \text{a.s. [P]}.$$

ii) If  $D_1 \in \{\rho_1, \mathcal{D}_1\}$ , and  $\tilde{E}(\mathcal{X})$  and  $D_1\text{-ADD}(\mathcal{X})$  exist, then the statistic  $\widehat{D_1\text{-ADD}}(\mathcal{X}_1, \dots, \mathcal{X}_n)$  is a strongly consistent estimator of  $D_1\text{-ADD}(\mathcal{X})$ , that is,

$$\lim_{n \rightarrow \infty} \widehat{D_1\text{-ADD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = D_1\text{-ADD}(\mathcal{X}) \quad a.s. [P].$$

iii) If for each  $\alpha \in [0, 1]$  the medians of the real-valued random variables  $\inf \mathcal{X}_\alpha$  and  $\sup \mathcal{X}_\alpha$  are unique, and the sequences of the real-valued sample medians  $\{\text{Me}\{\inf(\mathcal{X}_1)_\alpha, \dots, \inf(\mathcal{X}_n)_\alpha\}\}_n$  and  $\{\text{Me}\{\sup(\mathcal{X}_1)_\alpha, \dots, \sup(\mathcal{X}_n)_\alpha\}\}_n$  as functions of  $\alpha$  over  $[0, 1]$  are both uniformly integrable, then the statistic  $\widehat{\rho_1\text{-MDD}}(\mathcal{X}_1, \dots, \mathcal{X}_n)$  is a strongly consistent estimator of  $\rho_1\text{-MDD}(\mathcal{X})$ , that is,

$$\lim_{n \rightarrow \infty} \widehat{\rho_1\text{-MDD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = \rho_1\text{-MDD}(\mathcal{X}) \quad a.s. [P].$$

iv) If the population median of the real-valued random variable  $\text{wabl}(\mathcal{X})$  is unique, and also for each  $\alpha \in [0, 1]$  the population medians of the real-valued random variables  $\text{ldev}_{\mathcal{X}}(\alpha)$  and  $\text{rdev}_{\mathcal{X}}(\alpha)$  are actually unique, and the two sequences of sample medians  $\{\text{Me}\{\text{ldev}_{\mathcal{X}_1}(\alpha), \dots, \text{ldev}_{\mathcal{X}_n}(\alpha)\}\}_n$  and  $\{\text{Me}\{\text{rdev}_{\mathcal{X}_1}(\alpha), \dots, \text{rdev}_{\mathcal{X}_n}(\alpha)\}\}_n$  as functions of  $\alpha$  over  $[0, 1]$  are both uniformly integrable, then the statistic  $\widehat{\mathcal{D}_1\text{-MDD}}(\mathcal{X}_1, \dots, \mathcal{X}_n)$  is a strongly consistent estimator of  $\mathcal{D}_1\text{-MDD}(\mathcal{X})$ , that is,

$$\lim_{n \rightarrow \infty} \widehat{\mathcal{D}_1\text{-MDD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = \mathcal{D}_1\text{-MDD}(\mathcal{X}) \quad a.s. [P].$$

*Proof* i) If  $\bar{\mathcal{X}}_n$  denotes the sample mean estimator, then on the basis of the triangle inequality for  $\rho_2$  we have that for any  $i \in \{1, \dots, n\}$

$$\begin{aligned} \left| \rho_2(\mathcal{X}_i, \tilde{E}(\mathcal{X})) - \rho_2(\tilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \right| &\leq \rho_2(\mathcal{X}_i, \bar{\mathcal{X}}_n) \\ &\leq \rho_2(\mathcal{X}_i, \tilde{E}(\mathcal{X})) + \rho_2(\tilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \end{aligned}$$

and, hence, by taking squares on the three members of the inequality and averaging later over  $i$  we have that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left[ \rho_2(\mathcal{X}_i, \tilde{E}(\mathcal{X})) \right]^2 + \left[ \rho_2(\tilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \right]^2 - \rho_2(\tilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \frac{2}{n} \sum_{i=1}^n \rho_2(\mathcal{X}_i, \tilde{E}(\mathcal{X})) \\ \leq \left[ \widehat{\text{SD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) \right]^2 \leq \frac{1}{n} \sum_{i=1}^n \left[ \rho_2(\mathcal{X}_i, \tilde{E}(\mathcal{X})) \right]^2 + \left[ \rho_2(\tilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \right]^2 \\ + \rho_2(\tilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \frac{2}{n} \sum_{i=1}^n \rho_2(\mathcal{X}_i, \tilde{E}(\mathcal{X})) \end{aligned}$$

whence, by applying the Strong Law of Large Numbers for the real-valued random variables  $\rho_2(\mathcal{X}, \tilde{E}(\mathcal{X}))$  and  $\left[ \rho_2(\mathcal{X}, \tilde{E}(\mathcal{X})) \right]^2$  and the strong consistency of  $\bar{\mathcal{X}}_n$  in  $\rho_2$ -sense (as a consequence from Colubi *et al.* [4], in which an SLLN for random fuzzy sets has been obtained for the stronger metric of the supremum between fuzzy sets), we have that

$$\lim_{n \rightarrow \infty} \left[ \widehat{\text{SD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) \right]^2 = [\text{SD}(\mathcal{X})]^2 \quad a.s. [P],$$

and therefore

$$\lim_{n \rightarrow \infty} \widehat{\text{SD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = \text{SD}(\mathcal{X}) \quad a.s. [P].$$

- ii) On the basis of the triangle inequality for  $D_1$  we have that for any  $i \in \{1, \dots, n\}$

$$D_1(\mathcal{X}_i, \widetilde{E}(\mathcal{X})) - D_1(\widetilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n) \leq D_1(\mathcal{X}_i, \bar{\mathcal{X}}_n) \leq D_1(\mathcal{X}_i, \widetilde{E}(\mathcal{X})) + D_1(\widetilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n)$$

and, hence, by averaging over  $i$  we have that

$$\left| D_1\text{-}\widehat{\text{ADD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) - \frac{1}{n} \sum_{i=1}^n D_1(\mathcal{X}_i, \widetilde{E}(\mathcal{X})) \right| \leq D_1(\widetilde{E}(\mathcal{X}), \bar{\mathcal{X}}_n),$$

whence by applying the Strong Law of Large Numbers for the real-valued random variable  $D_1(\mathcal{X}, \widetilde{E}(\mathcal{X}))$  and the strong consistency of  $\bar{\mathcal{X}}_n$  in  $D_1$ -sense we have that

$$\left| \lim_{n \rightarrow \infty} D_1\text{-}\widehat{\text{ADD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) - D_1\text{-}\text{ADD}(\mathcal{X}) \right| = 0 \quad a.s. [P],$$

what implies that

$$\lim_{n \rightarrow \infty} D_1\text{-}\widehat{\text{ADD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = D_1\text{-}\text{ADD}(\mathcal{X}) \quad a.s. [P].$$

- iii) On the basis of the triangle inequality for  $\rho_1$  for any  $i \in \{1, \dots, n\}$  we have that

$$\begin{aligned} \rho_1(\mathcal{X}_i, \widetilde{\text{Me}}(\mathcal{X})) - \rho_1(\widetilde{\text{Me}}(\mathcal{X}), \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)) &\leq \rho_1(\mathcal{X}_i, \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)) \\ &\leq \rho_1(\mathcal{X}_i, \widetilde{\text{Me}}(\mathcal{X})) + \rho_1(\widetilde{\text{Me}}(\mathcal{X}), \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)) \end{aligned}$$

and, hence,

$$\begin{aligned} \text{Me}\{\rho_1(\mathcal{X}_1, \widetilde{\text{Me}}(\mathcal{X})), \dots, \rho_1(\mathcal{X}_n, \widetilde{\text{Me}}(\mathcal{X}))\} - \rho_1(\widetilde{\text{Me}}(\mathcal{X}), \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)) \\ \leq \rho_1\text{-}\widehat{\text{MDD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) \\ \leq \text{Me}\{\rho_1(\mathcal{X}_1, \widetilde{\text{Me}}(\mathcal{X})), \dots, \rho_1(\mathcal{X}_n, \widetilde{\text{Me}}(\mathcal{X}))\} + \rho_1(\widetilde{\text{Me}}(\mathcal{X}), \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)) \end{aligned}$$

whence, by applying the strong consistency of the fuzzy-valued sample 1-norm median in  $\rho_1$ -sense (see Sinova *et al.* [26]) and the strong consistency under the assumed conditions of the sample median from the real-valued random variable  $\rho_1(\mathcal{X}, \widetilde{\text{Me}}(\mathcal{X}))$ , we have that

$$\lim_{n \rightarrow \infty} \rho_1\text{-}\widehat{\text{MDD}}(\mathcal{X}_1, \dots, \mathcal{X}_n) = \rho_1\text{-}\text{MDD}(\mathcal{X}) \quad a.s. [P].$$

- iv) The proof is analogous to that for *iii*) by applying the strong consistency of the fuzzy-valued sample wabl/ldev/rdev median in  $\mathcal{D}_1$ -sense (see Sinova *et al.* [27]).  $\square$

*Remark 4* It should be pointed out that the strong consistency in Proposition 8 has been understood in the sense of the convergence of the sample scale measure to the population one, instead of the convergence (after maybe a correction) of the sample measure to the population standard deviation as often made in case of dealing with real-valued random variables. Actually, the last convergence is usually discussed in the real-valued case by considering some outstanding models for the distribution of the involved random variables; this type of models has not been yet stated for RFNs in a realistic way.

#### 4 Analysis of the robustness: the finite sample breakdown point

A popular and powerful tool allowing us to describe the robustness of an estimator is its breakdown point. As we can read in Donoho and Huber [9], “the notion of breakdown point was coined, formally defined, and very briefly discussed by Frank Hampel, at that time a student of Erich Lehman, in his PhD in 1968” [13]. Although originally it was presented for location estimators, the concept has also been generalized to scale estimators.

A simple and intuitive definition of the breakdown point but one restricted to finite samples, the so-called **finite sample breakdown point** (fsbp for short), was introduced by Donoho [8] and Donoho and Huber [9]. For scale estimators, it is defined as the minimum proportion of sample data which should be perturbed in order to let the estimator get either an arbitrary large value or the value zero. The higher the breakdown point of an estimator, the more robust it is. Therefore, two situations are required to study: the one consisting in contaminating the sample by means of outliers, which can make the estimator overestimate the true scale up to infinity (explosion), and the one consisting in contaminating the sample by means of inliers, which may result in underestimation of the true scale to zero (implosion). Notice that when we deal with location estimators only the explosion case needs to be considered.

Next, the replacement version of the finite sample breakdown point for scale estimators (see Donoho and Huber [9]) is recalled (applied in this case to imprecise data):

**Definition 13** For any sample of observations  $\tilde{\mathbf{x}}_n$  from an RFN  $\mathcal{X}$ , the finite sample breakdown point of a scale estimate  $\hat{T}_n$  is defined by

$$\text{fsbp}^*(\hat{T}_n, \tilde{\mathbf{x}}_n) = \min \left\{ \text{fsbp}^+(\hat{T}_n, \tilde{\mathbf{x}}_n), \text{fsbp}^-(\hat{T}_n, \tilde{\mathbf{x}}_n) \right\}$$

where

$$\text{fsbp}^+(\hat{T}_n, \tilde{\mathbf{x}}_n) = \min \left\{ \frac{k}{n}; \sup_{\tilde{\mathbf{y}}_{n,k}} \hat{T}(\tilde{\mathbf{y}}_{n,k}) = \infty \right\}$$

and

$$\text{fsbp}^-(\hat{T}_n, \tilde{\mathbf{x}}_n) = \min \left\{ \frac{k}{n}; \inf_{\tilde{\mathbf{y}}_{n,k}} \hat{T}(\tilde{\mathbf{y}}_{n,k}) = 0 \right\}$$

with  $\tilde{\mathbf{y}}_{n,k}$  obtained by replacing any  $k$  observations of  $\tilde{\mathbf{x}}_n$  by arbitrary values. The quantities  $\text{fsbp}^+$  and  $\text{fsbp}^-$  are called the **explosion** breakdown point and the **implosion** breakdown point.

#### 4.1 General analysis of the finite sample breakdown point

In the following theorems it will be proved that if in the considered sample  $\tilde{\mathbf{x}}_n$  there are no equal observations, then the fsbp of the estimate  $\widehat{D_1\text{-MDD}}(\tilde{\mathbf{x}}_n)$  is  $\frac{1}{n} \lfloor \frac{n}{2} \rfloor$  (denoting  $\lfloor \cdot \rfloor$  the floor function), which is the highest possible fsbp for a scale estimate. By contrast, the fsbp of the estimates  $\widehat{SD}(\tilde{\mathbf{x}}_n)$  and  $\widehat{D_1\text{-ADD}}(\tilde{\mathbf{x}}_n)$  is  $\frac{1}{n}$ , the lowest possible. The proofs are based on Rousseeuw and Croux [25].

**Theorem 1** *For any sample  $\tilde{\mathbf{x}}_n = \{\tilde{x}_1, \dots, \tilde{x}_n\}$  from an RFN  $\mathcal{X}$  in which there are not two identical fuzzy numbers, we have that*

$$\text{fsbp}^+(\widehat{D_1\text{-ADD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n}, \quad \text{fsbp}^-(\widehat{D_1\text{-ADD}}_n, \tilde{\mathbf{x}}_n) = \frac{n-1}{n}.$$

Therefore, the finite sample breakdown point of the scale estimate  $\widehat{D_1\text{-ADD}}_n$  is given by

$$\text{fsbp}^*(\widehat{D_1\text{-ADD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n}.$$

*Proof* Let  $\tilde{\mathbf{x}}_n = \{\tilde{x}_1, \dots, \tilde{x}_n\}$  be a sample in which there are not two identical fuzzy numbers, and denote  $\text{fsbp}^+ = \text{fsbp}^+(\widehat{D_1\text{-ADD}}_n, \tilde{\mathbf{x}}_n)$  and  $\text{fsbp}^- = \text{fsbp}^-(\widehat{D_1\text{-ADD}}_n, \tilde{\mathbf{x}}_n)$ . The proof is to be split in three steps.

*Step 1:* We begin showing that  $\text{fsbp}^- \leq (n-1)/n$ .

We are going to find a sample  $\tilde{\mathbf{y}}_{n,k} = \{\tilde{y}_1, \dots, \tilde{y}_n\}$  with  $k = n-1$  replaced observations such that  $\widehat{D_1\text{-ADD}}(\tilde{\mathbf{y}}_{n,k}) = 0$ .

We construct the sample  $\tilde{\mathbf{y}}_{n,k}$  by replacing the observations  $\tilde{x}_2, \dots, \tilde{x}_n$  by  $\tilde{x}_1$ . The considered sample  $\tilde{\mathbf{y}}_{n,k}$  has  $k = n-1$  replaced observations. See that  $\widehat{D_1\text{-ADD}}(\tilde{\mathbf{y}}_{n,n-1}) = 0$ .

Since  $\tilde{\mathbf{y}}_{n,n-1} = \tilde{x}_1$ , we have that for all  $i \in \{1, \dots, n\}$ ,  $D_1(\tilde{y}_i, \tilde{\mathbf{y}}_{n,n-1}) = D_1(\tilde{x}_1, \tilde{x}_1) = 0$

Thus,  $\widehat{D_1\text{-ADD}}(\tilde{\mathbf{y}}_{n,n-1}) = \sum_{i=1}^n D_1(\tilde{y}_i, \tilde{\mathbf{y}}_{n,n-1})/n = 0$ .

Therefore,  $\inf_{\tilde{\mathbf{z}}_{n,n-1}} \widehat{D_1\text{-ADD}}(\tilde{\mathbf{z}}_{n,n-1}) = 0$  with  $\tilde{\mathbf{z}}_{n,n-1}$  any sample with  $n-1$  replaced observations, and for this reason  $\text{fsbp}^- \leq (n-1)/n$ .

*Step 2:* Now we show that  $\text{fsbp}^- \geq (n-1)/n$ .

Let  $\tilde{\mathbf{y}}_{n,k}$  be any sample with  $k < n-1$  replaced observations. There exist at least two observations  $\tilde{x}_1^{\tilde{\mathbf{y}}_{n,k}}, \tilde{x}_2^{\tilde{\mathbf{y}}_{n,k}} \in \{\tilde{x}_1, \dots, \tilde{x}_n\}$  such that  $\tilde{x}_1^{\tilde{\mathbf{y}}_{n,k}}, \tilde{x}_2^{\tilde{\mathbf{y}}_{n,k}} \in \tilde{\mathbf{y}}_{n,k}$ . Let  $\delta := \min_{i,j \in \{1, \dots, n\}} D_1(\tilde{x}_i, \tilde{x}_j)/n > 0$ . We have that

$$\begin{aligned} \delta &= \frac{\min_{i,j \in \{1, \dots, n\}} D_1(\tilde{x}_i, \tilde{x}_j)}{n} \leq \frac{D_1(\tilde{x}_1^{\tilde{\mathbf{y}}_{n,k}}, \tilde{x}_2^{\tilde{\mathbf{y}}_{n,k}})}{n} \\ &\leq \frac{D_1(\tilde{x}_1^{\tilde{\mathbf{y}}_{n,k}}, \tilde{\mathbf{y}}_{n,k}) + D_1(\tilde{x}_2^{\tilde{\mathbf{y}}_{n,k}}, \tilde{\mathbf{y}}_{n,k})}{n} \leq \frac{\sum_{i=1}^n D_1(\tilde{y}_i, \tilde{\mathbf{y}}_{n,k})}{n} = \widehat{D_1\text{-ADD}}(\tilde{\mathbf{y}}_{n,k}) \end{aligned}$$

Therefore,  $\inf_{\tilde{\mathbf{y}}_{n,k}} \widehat{D_1\text{-ADD}}(\tilde{\mathbf{y}}_{n,k}) \geq \delta > 0$  with  $k < n-1$ , and so,  $\text{fsbp}^- \geq (n-1)/n$ .

*Step 3:* Finally, we will prove that  $\text{fsbp}^+ = 1/n$ .

We construct the sample  $\tilde{\mathbf{y}}_{n,k}$  by replacing the observation  $\tilde{x}_1$  by  $\tilde{x}_{(n)} + L$ , with  $L \in \mathbb{R}$ ,  $L > 0$  and  $\tilde{x}_{(n)}$  the fuzzy number defined for each  $\alpha \in [0, 1]$  as follows:

$$(\tilde{x}_{(n)})_\alpha = \left[ \max\{\inf(\tilde{x}_i)_\alpha : i = 1, \dots, n\}, \max\{\sup(\tilde{x}_i)_\alpha : i = 1, \dots, n\} \right].$$

Indeed,  $\tilde{x}_{(n)} \in \mathcal{F}_c^*(\mathbb{R})$  since by considering its inf/sup representation/characterization (see, for instance, Goetschel and Voxman [11]) one immediately verifies that

- i)*  $-\inf(\tilde{x}_{(n)})_\alpha$  and  $\sup(\tilde{x}_{(n)})_\alpha$  are left-continuous on  $(0, 1]$ , right-continuous at 0, and non-increasing on  $[0, 1]$ ,
- ii)* and  $\inf(\tilde{x}_{(n)})_1 \leq \sup(\tilde{x}_{(n)})_1$ .

The considered sample  $\tilde{\mathbf{y}}_{n,k}$  has  $k = 1$  replaced observation. In this case, it satisfies that

$$\begin{aligned} D_1(\tilde{y}_1, \tilde{\mathbf{y}}_{n,1}) &= D_1(\tilde{x}_{(n)} + L, \tilde{\mathbf{y}}_{n,1}) = D_1\left(\tilde{x}_{(n)} + L, \tilde{\mathbf{x}}'_n + \frac{L}{n}\right) \\ &= D_1\left(\tilde{x}_{(n)}, \tilde{\mathbf{x}}'_n\right) + D_1\left(L, \frac{L}{n}\right) = D_1\left(\tilde{x}_{(n)}, \tilde{\mathbf{x}}'_n\right) + \frac{(n-1)L}{n} \geq \frac{(n-1)L}{n}, \end{aligned}$$

with  $\tilde{\mathbf{x}}'_n = \{\tilde{x}_{(n)}, \tilde{x}_2, \dots, \tilde{x}_n\}$ , and taking into account in the third equality that for all  $\alpha \in [0, 1]$ ,  $\inf(\tilde{x}_{(n)})_\alpha \geq \inf(\tilde{\mathbf{x}}'_n)_\alpha$ ,  $\sup(\tilde{x}_{(n)})_\alpha \geq \sup(\tilde{\mathbf{x}}'_n)_\alpha$  and  $L \geq L/n$ .

Thus,  $D_{1-\text{ADD}}(\tilde{\mathbf{y}}_{n,1}) = \sum_{i=1}^n D_1(\tilde{y}_i, \tilde{\mathbf{y}}_{n,1})/n \geq (n-1)L/n^2$ .

Letting  $L \rightarrow \infty$ ,  $\sup_{\tilde{\mathbf{z}}_{n,1}} D_{1-\text{ADD}}(\tilde{\mathbf{z}}_{n,1}) = \infty$  with  $\tilde{\mathbf{z}}_{n,1}$  any sample with 1 replaced observations. Therefore

$$\text{fsbp}^+ = \frac{1}{n}. \quad \square$$

**Theorem 2** For any sample of observations  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  from an RFN  $\mathcal{X}$  in which there are not two identical fuzzy numbers, we have that

$$\text{fsbp}^+(\widehat{\text{SD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n}, \quad \text{fsbp}^-(\widehat{\text{SD}}_n, \tilde{\mathbf{x}}_n) = \frac{n-1}{n}.$$

Therefore, the finite sample breakdown point of the scale estimate  $\widehat{\text{SD}}_n$  is given by

$$\text{fsbp}^*(\widehat{\text{SD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n}.$$

*Proof* It is analogous to the previous one. □

**Theorem 3** For any sample of observations  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  from an RFN  $\mathcal{X}$  in which there are not two identical fuzzy numbers, we have that

$$\text{fsbp}^+(\rho_1\text{-MDD}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor, \quad \text{fsbp}^-(\rho_1\text{-MDD}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n} \left\lfloor \frac{n}{2} \right\rfloor.$$

Therefore, the finite sample breakdown point of the scale estimate  $\rho_1\text{-MDD}_n$  is given by

$$\text{fsbp}^*(\rho_1\text{-MDD}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n} \left\lfloor \frac{n}{2} \right\rfloor.$$

*Proof* Denote  $\text{fsbp}^+ = \text{fsbp}^+(\rho_1\text{-MDD}_n, \tilde{\mathbf{x}}_n)$  and  $\text{fsbp}^- = \text{fsbp}^-(\rho_1\text{-MDD}_n, \tilde{\mathbf{x}}_n)$ . The proof is presented in four steps.

*Step 1:* We begin showing that  $\text{fsbp}^- \leq \lfloor \frac{n}{2} \rfloor / n$ .

We are going to find a sample  $\tilde{\mathbf{y}}_{n,k} = \{\tilde{y}_1, \dots, \tilde{y}_n\}$  with  $k = \lfloor \frac{n}{2} \rfloor$  replaced observations such that  $\rho_1\text{-MDD}(\tilde{\mathbf{y}}_{n,k}) = 0$ .

We construct the sample  $\tilde{\mathbf{y}}_{n,k}$  by replacing the observations  $\tilde{x}_2, \dots, \tilde{x}_{\lfloor \frac{n}{2} \rfloor + 1}$  by  $\tilde{x}_1$ . The sample  $\tilde{\mathbf{y}}_{n,k}$  has  $k = \lfloor \frac{n}{2} \rfloor$  replaced observations. See that  $\rho_1\text{-MDD}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n}{2} \rfloor}) = 0$ . Since  $\widehat{\text{Me}}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n}{2} \rfloor}) = \tilde{x}_1$ , then for all  $i \in \{1, \dots, \lfloor \frac{n}{2} \rfloor + 1\}$ ,

$$\rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n}{2} \rfloor})) = \rho_1(\tilde{x}_1, \tilde{x}_1) = 0.$$

$$\text{Thus, } \rho_1\text{-MDD}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n}{2} \rfloor}) = \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n}{2} \rfloor})) \right) = 0.$$

Therefore,  $\inf_{\tilde{\mathbf{z}}_{n, \lfloor \frac{n}{2} \rfloor}} \rho_1\text{-MDD}(\tilde{\mathbf{z}}_{n, \lfloor \frac{n}{2} \rfloor}) = 0$  for any sample  $\tilde{\mathbf{z}}_{n, \lfloor \frac{n}{2} \rfloor}$  with  $\lfloor \frac{n}{2} \rfloor$  replaced observations, and  $\text{fsbp}^- \leq \lfloor \frac{n}{2} \rfloor / n$ .

*Step 2:* Now we show that  $\text{fsbp}^- \geq \lfloor \frac{n}{2} \rfloor / n$ .

Take any sample  $\tilde{\mathbf{y}}_{n,k}$  with  $k < \lfloor \frac{n}{2} \rfloor$  replaced observations. Because of the definition of median for real numbers, there exist at least  $\lfloor \frac{n}{2} \rfloor + 1$  observations  $\tilde{y}^j$  (with  $j \in \{1, \dots, \lfloor \frac{n}{2} \rfloor + 1\}$ ) in the sample  $\tilde{\mathbf{y}}_{n,k}$  such that  $\rho_1(\tilde{y}^j, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) / 2 \leq \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \right)$ .

Moreover, because of  $\tilde{\mathbf{y}}_{n,k}$  having  $k < \lfloor \frac{n}{2} \rfloor$  replaced observations from the sample  $\tilde{\mathbf{x}}_n$ , there exist at least two observations  $\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1, \tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^2 \in \{\tilde{x}_1, \dots, \tilde{x}_n\}$  such that  $\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1 = \tilde{y}^{j_1}$  and  $\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^2 = \tilde{y}^{j_2}$  with  $\tilde{y}^{j_1}, \tilde{y}^{j_2} \in \{\tilde{y}^1, \dots, \tilde{y}^{\lfloor \frac{n}{2} \rfloor + 1}\}$ . Let  $\delta := \min_{i,j \in \{1, \dots, n\}} \rho_1(\tilde{x}_i, \tilde{x}_j) / 4 > 0$ . We have that

$$\begin{aligned} \delta &= \frac{\min_{i,j \in \{1, \dots, n\}} \rho_1(\tilde{x}_i, \tilde{x}_j)}{4} \leq \frac{\rho_1(\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1, \tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^2)}{4} \\ &\leq \frac{1}{2} \cdot \frac{\rho_1(\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) + \rho_1(\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^2, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k}))}{2} \\ &\leq \frac{1}{2} \cdot \left( \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \right) + \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \right) \right) \end{aligned}$$

$$= \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \right) = \rho_1\text{-}\widehat{\text{MDD}}(\tilde{\mathbf{y}}_{n,k}).$$

Therefore,  $\inf_{\tilde{\mathbf{y}}_{n,k}} \rho_1\text{-}\widehat{\text{MDD}}(\tilde{\mathbf{y}}_{n,k}) \geq \delta > 0$  with  $k < \lfloor \frac{n}{2} \rfloor$ , and hence  $\text{fsbp}^- \geq \lfloor \frac{n}{2} \rfloor / n$ .

*Step 3:* Now we will prove that  $\text{fsbp}^+ \leq \lfloor \frac{n+1}{2} \rfloor / n$ .

We construct the sample  $\tilde{\mathbf{y}}_{n,k}$  by replacing the observation  $\tilde{x}_1$  by  $\tilde{x}_{(n)} + L$ ,  $\tilde{x}_2$  by  $\tilde{x}_{(n)} + 2L$ ,  $\dots$ ,  $\tilde{x}_{\lfloor \frac{n+1}{2} \rfloor}$  by  $\tilde{x}_{(n)} + \lfloor \frac{n+1}{2} \rfloor L$ , with  $L \in \mathbb{R}$ ,  $L > 0$  and  $\tilde{x}_{(n)}$  the fuzzy number defined for each  $\alpha \in [0, 1]$  by

$$(\tilde{x}_{(n)})_\alpha = \left[ \max\{\inf(\tilde{x}_i)_\alpha : i = 1, \dots, n\}, \max\{\sup(\tilde{x}_i)_\alpha : i = 1, \dots, n\} \right].$$

The considered sample  $\tilde{\mathbf{y}}_{n,k}$  has  $k = \lfloor \frac{n+1}{2} \rfloor$  replaced observations. It satisfies that  $\rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \geq L/2$  for any  $i \in \{2, \dots, n\}$ .

Hence,  $\rho_1\text{-}\widehat{\text{MDD}}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n+1}{2} \rfloor}) = \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n, \lfloor \frac{n+1}{2} \rfloor})) \right) \geq L/2$ . Consequently, as  $L \rightarrow \infty$ , we have that  $\sup_{\tilde{\mathbf{z}}_{n, \lfloor \frac{n+1}{2} \rfloor}} \rho_1\text{-}\widehat{\text{MDD}}(\tilde{\mathbf{z}}_{n, \lfloor \frac{n+1}{2} \rfloor}) = \infty$  for any sample  $\tilde{\mathbf{z}}_{n, \lfloor \frac{n+1}{2} \rfloor}$  with  $\lfloor \frac{n+1}{2} \rfloor$  replaced observations.

*Step 4:* Finally, we will prove that  $\text{fsbp}^+ \geq \lfloor \frac{n+1}{2} \rfloor / n$ .

Take any sample  $\tilde{\mathbf{y}}_{n,k}$  with  $k < \lfloor \frac{n+1}{2} \rfloor$  replaced observations. Because of the definition of median for real numbers, there exist at least  $\lfloor \frac{n+1}{2} \rfloor$  observations  $\tilde{y}^j$  of the sample  $\tilde{\mathbf{y}}_{n,k}$  such that  $\rho_1(\tilde{y}^j, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \geq \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \right)$ , with  $j \in \{1, \dots, \lfloor \frac{n+1}{2} \rfloor\}$ .

Moreover, because of  $\tilde{\mathbf{y}}_{n,k}$  having  $k < \lfloor \frac{n+1}{2} \rfloor$  replaced observations from the sample  $\tilde{\mathbf{x}}_n$ , there exist at least one observation  $\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1 \in \{\tilde{x}_1, \dots, \tilde{x}_n\}$  such that  $\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1 = \tilde{y}^j$  with  $\tilde{y}^j \in \{\tilde{y}^1, \dots, \tilde{y}^{\lfloor \frac{n+1}{2} \rfloor}\}$ . Let  $M := \max_{i,j \in \{1, \dots, n\}} \rho_1(\tilde{x}_i, \tilde{x}_j) < \infty$ . We have that

$$\begin{aligned} M &= \max_{i,j \in \{1, \dots, n\}} \rho_1(\tilde{x}_i, \tilde{x}_j) \geq \rho_1(\tilde{x}_{\tilde{\mathbf{y}}_{n,k}}^1, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \\ &\geq \text{Me}_{i \in \{1, \dots, n\}} \left( \rho_1(\tilde{y}_i, \widehat{\text{Me}}(\tilde{\mathbf{y}}_{n,k})) \right) = \rho_1\text{-}\widehat{\text{MDD}}(\tilde{\mathbf{y}}_{n,k}). \end{aligned}$$

Therefore,  $\sup_{\tilde{\mathbf{y}}_{n,k}} \rho_1\text{-}\widehat{\text{MDD}}(\tilde{\mathbf{y}}_{n,k}) \leq M < \infty$  with  $k < \lfloor \frac{n+1}{2} \rfloor$ , and consequently,  $\text{fsbp}^+ \geq \lfloor \frac{n+1}{2} \rfloor / n$ .  $\square$

**Theorem 4** For any sample of observations  $\tilde{\mathbf{x}}_n = (\tilde{x}_1, \dots, \tilde{x}_n)$  from an RFN  $\mathcal{X}$  in which there are not two identical observations, we have that

$$\text{fsbp}^+(\mathcal{D}_1\text{-}\widehat{\text{MDD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor, \quad \text{fsbp}^-(\mathcal{D}_1\text{-}\widehat{\text{MDD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n} \left\lfloor \frac{n}{2} \right\rfloor.$$

Therefore, the finite sample breakdown point of the scale estimate  $\mathcal{D}_1\text{-}\widehat{\text{MDD}}_n$  is given by

$$\text{fsbp}^*(\mathcal{D}_1\text{-}\widehat{\text{MDD}}_n, \tilde{\mathbf{x}}_n) = \frac{1}{n} \left\lfloor \frac{n}{2} \right\rfloor.$$

*Proof* It is analogous to the previous one.  $\square$

#### 4.2 Simulations-based analysis of the finite sample breakdown point: empirical bias curves

To illustrate in which way the contamination affects the considered scale estimators, a simulation study is now presented.

Following Sinova *et al.* [26], trapezoidal RFNs  $\mathcal{X} = \text{Tra}(\inf \mathcal{X}_0, \inf \mathcal{X}_1, \sup \mathcal{X}_1, \sup \mathcal{X}_0)$  have been considered in the simulations, each of them characterized by means of the following four real-valued random variables, namely,

$$X_1 = \text{mid}\mathcal{X}_1 = (\inf \mathcal{X}_1 + \sup \mathcal{X}_1)/2, \quad X_2 = \text{spr}\mathcal{X}_1 = (\sup \mathcal{X}_1 - \inf \mathcal{X}_1)/2,$$

$$X_3 = \text{lspr}\mathcal{X}_0 = \inf \mathcal{X}_1 - \inf \mathcal{X}_0, \quad X_4 = \text{rspr}\mathcal{X}_0 = \sup \mathcal{X}_0 - \sup \mathcal{X}_1,$$

whence  $\mathcal{X} = \text{Tra}(X_1 - X_2 - X_3, X_1 - X_2, X_1 + X_2, X_1 + X_2 + X_4)$ .

Recall that when we deal with scale estimators, two different types of contamination need to be studied: the one caused by the presence of outliers in the sample, which can make the estimator explode to infinite (explosion), and the one caused by the presence of inliers in the sample, which can make the estimator implode to zero (implosion).

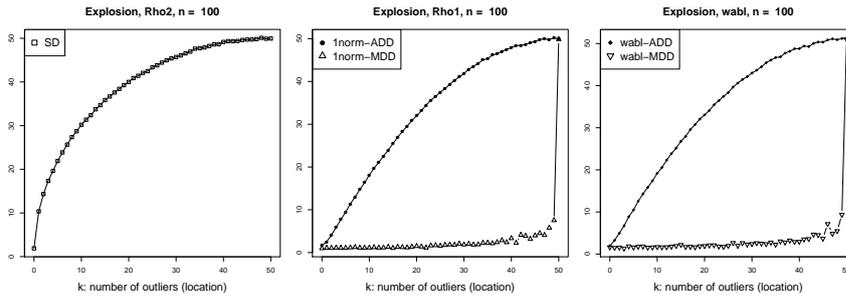
##### **Explosion case:**

Three different types of outliers have been considered:

- Outliers because of the location:  $X_1 \sim \mathcal{N}(100, 1)$  and  $X_2, X_3, X_4 \sim \chi_1^2$ .
- Outliers because of the shape:  $X_1 \sim \mathcal{N}(0, 1)$  and  $X_2, X_3, X_4 \sim \chi_1^2 + 100$ .
- Outliers because of the location and shape:  $X_1 \sim \mathcal{N}(100, 1)$  and  $X_2, X_3, X_4 \sim \chi_1^2 + 100$ .

For each type of outlier, the next steps have been followed:

- A sample of size  $n = 100$  of trapezoidal fuzzy numbers has been generated, assuming that  $X_1 \sim \mathcal{N}(0, 1)$ ,  $X_2, X_3, X_4 \sim \chi_1^2$ , being all of them independent.
- $k$  observations of the sample have been replaced by the outliers, with  $k \in \{0, 1, \dots, 50\}$ .
- For each  $k$ , the value of the different scale estimators has been calculated.



**Fig. 1** Explosion breakdown point with outliers because of the location

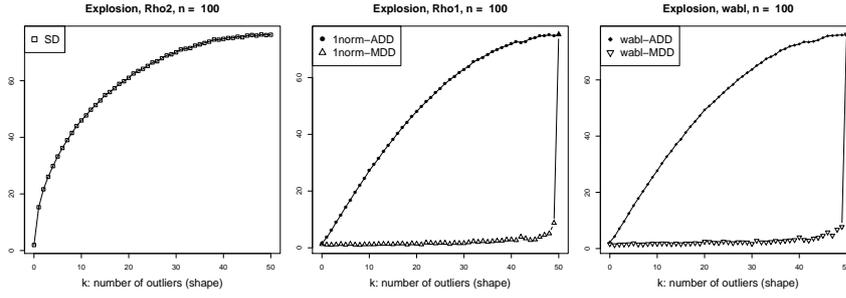


Fig. 2 Explosion breakdown point with outliers because of the shape

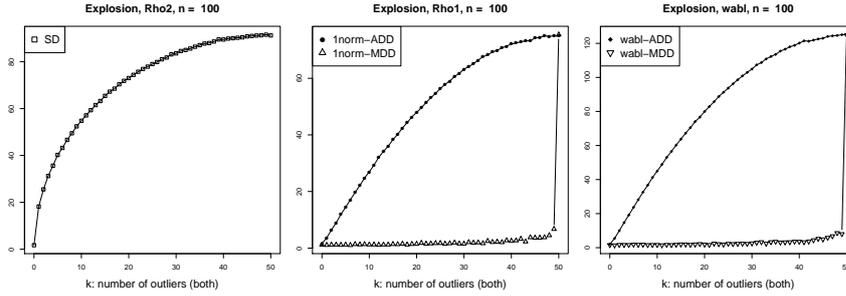


Fig. 3 Explosion breakdown point with outliers because of the location and shape

Figures 1, 2 and 3 display the explosion bias curves with the three different types of outliers. As we can check, the breakdown for the estimators  $\widehat{SD}_n$  and  $\widehat{D_1-ADD}_n$  happens when a single outlier is introduced in the sample (the value of these estimators changes considerably from 0 to 1 outlier). However, 50 outliers are needed to introduce to cause breakdown with the estimators  $\widehat{D_1-MDD}_n$ .

#### Implosion case:

Inliers must be considered as contamination in the sample:

- A sample of size  $n = 100$  of trapezoidal fuzzy numbers has been generated, assuming that  $X_1 \sim \mathcal{N}(0, 1)$ ,  $X_2, X_3, X_4 \sim \chi_1^2$ , being all of them independent.
- One datum of the sample has been randomly chosen and  $k$  observations have been replaced by it, with  $k \in \{0, 1, \dots, 99\}$ .
- For each  $k$ , the value of the different scale estimators has been computed.

On the basis of the simulations for the implosion case one can empirically conclude from Figure 4, that 99 observations are needed to perturb to let the estimators  $\widehat{SD}_n$  and  $\widehat{D_1-ADD}_n$  implode to the value zero. By contrast, only 50 are needed to contaminate for the estimators  $\widehat{D_1-MDD}_n$ .

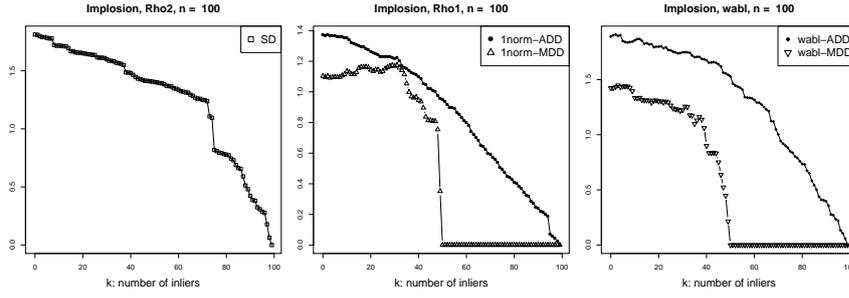


Fig. 4 Implosion breakdown point

## 5 Analysis of robustness: Sensitivity curves

In the previous section, the empirical bias curves have been studied as a tool to measure the robustness of an estimator. Now, another important and useful measurement of the robustness will be considered: the sensitivity curves, which are the sample version of the influence functions (see Maronna *et al.* [19]).

Whereas the empirical bias curves tell us how much the estimator changes when a percentage of the data is contaminated, the sensitivity curves describe how the estimator reacts to a single outlier.

Dealing with fuzzy data, we can consider three different types of outliers depending on if they deviate from the majority of data because of their location, because of their shape or for both reasons. Therefore, some sensitivity curves allow us to measure the effect of different locations of a single outlier in the sample, other curves measure the effect of considering different shapes of a single outlier, and other ones measure the effect of different locations and shapes of a single outlier in the sample.

Following the simulation procedure performed in the previous section, for each type of outlier, a sample of size  $n = 100$  of trapezoidal fuzzy numbers,  $\tilde{\mathbf{x}}_{100} = \{\tilde{x}_1, \dots, \tilde{x}_{100}\}$ , was considered, assuming that  $X_1 \sim \mathcal{N}(0, 1)$ ,  $X_2, X_3, X_4 \sim \chi_1^2$ , being all of them independent. The outlier  $\tilde{x}_a$  was constructed by replacement in the following way:

Let  $\tilde{x}_i \in \{\tilde{x}_1, \dots, \tilde{x}_{100}\}$  be characterized by the four-tuple  $(x_i^1, x_i^2, x_i^3, x_i^4)$ , with  $x_i^1 \in X_1$ ,  $x_i^2 \in X_2$ ,  $x_i^3 \in X_3$  and  $x_i^4 \in X_4$ , the arbitrary observation that will be replaced by the outlier  $\tilde{x}_a$ . Depending on the type of outlier,  $\tilde{x}_a \sim (x_a^1, x_a^2, x_a^3, x_a^4)$  was constructed as follows

- Outlier because of the location:  $x_a^1 = x_i^1 + a$ ,  $x_a^2 = x_i^2$ ,  $x_a^3 = x_i^3$ ,  $x_a^4 = x_i^4$ , with  $a \in \mathbb{R}$ .
- Outlier because of the shape:  $x_a^1 = x_i^1$ ,  $x_a^2 = x_i^2 + |a|$ ,  $x_a^3 = x_i^3 + |a|$ ,  $x_a^4 = x_i^4 + |a|$ , with  $a \in \mathbb{R}$ .
- Outlier because of the location and shape:  $x_a^1 = x_i^1 + a$ ,  $x_a^2 = x_i^2 + |a|$ ,  $x_a^3 = x_i^3 + |a|$ ,  $x_a^4 = x_i^4 + |a|$ , with  $a \in \mathbb{R}$ .

**Definition 14** Let  $\mathcal{X}$  be an RFN,  $\tilde{\mathbf{x}}_n = \{\tilde{x}_1, \dots, \tilde{x}_n\}$  a sample of observations and  $\hat{T}_n$  a scale estimator. The sensitivity curve of the estimator  $\hat{T}_n$  for the

sample  $\tilde{\mathbf{x}}_n$  is the difference

$$SC(\tilde{x}_a) = \widehat{T}(\tilde{\mathbf{x}}'_n) - \widehat{T}(\tilde{\mathbf{x}}_n)$$

where the sample  $\tilde{\mathbf{x}}'_n$  is obtained by replacing any observation of  $\tilde{\mathbf{x}}_n$  by the outlier  $\tilde{x}_a$ , with  $a \in \mathbb{R}$ .

The sensitivity curve of a robust scale estimator will be bounded, while the one corresponding to a non-robust scale estimator will have an unbounded behaviour.

For the simulations, the three types of outliers were considered and the real number  $a$  varying in  $[-10, 10]$ . Figure 5, 6 and 7 plot the sensitivity curves of the different scale estimators, for each type of outlier.

In all cases, we can see that the sensitivity curves with estimators  $\widehat{D}_1\text{-MDD}_n$  are bounded, whereas the ones corresponding to the estimators  $\widehat{SD}_n$  and  $\widehat{D}_1\text{-ADD}_n$  have an unbounded behaviour.

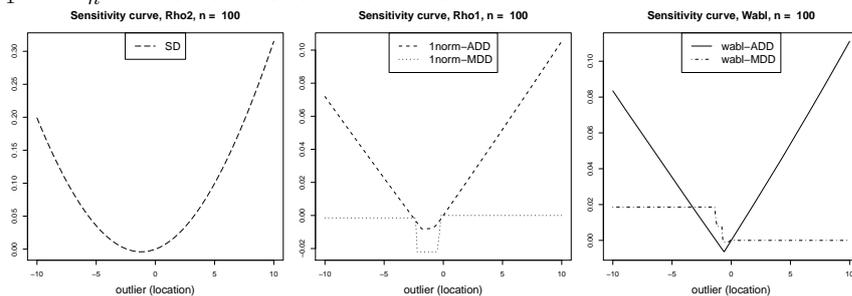


Fig. 5 Sensitivity curves with outliers because of the location

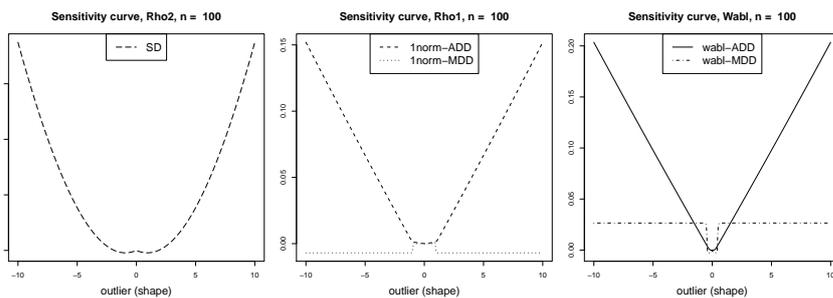


Fig. 6 Sensitivity curves with outliers because of the shape

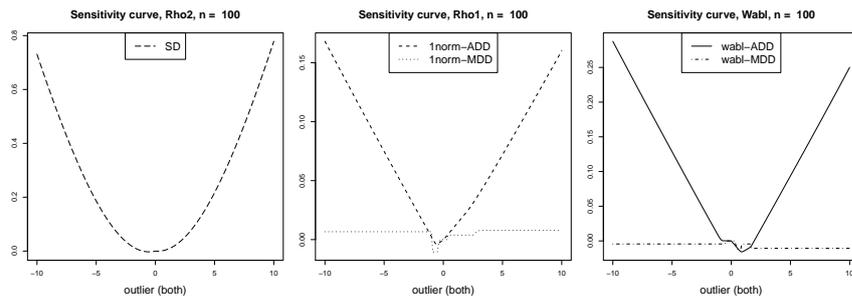


Fig. 7 Sensitivity curves with outliers because of the location and shape

## 6 Concluding remarks

This paper presents an introductory work to the analysis of the robustness of scale measures in dealing with imprecise data, when they are assumed to be modelled by means of fuzzy numbers and generated from random fuzzy numbers. Several scale estimators have been presented and some of their properties have been examined.

The finite sample breakdown point, the empirical bias curves and the sensitivity curves, as powerful tools to verify the robust behaviour of the considered estimators, have allowed us to draw similar conclusions to those from the case of real-valued data.

The study conducted in this article can be complemented and extended in the future by facing, among others,

- the formalization and discussion of other properties of the scale estimators,
- the use of other metrics between fuzzy data,
- the study of other scale estimates, like the indices  $\widehat{S}_n$  and  $\widehat{Q}_n$  introduced by Rousseeuw and Croux (see [25]), and the M-estimates of scale,
- the use of other tools to check the robustness.

## References

1. Blanco-Fernandez, A., Casals, M.R., Colubi, A., Corral, N., Garcıa-Barzana, M., Gil, M.A., Gonzalez-Rodrıguez, G., Lopez, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., de la Rosa de Saa, S., Sinova, B.: A distance-based statistical analysis of fuzzy number-valued data. *Int. J. Approx. Reas.* 55, 1487–1501 (2014)
2. Blanco-Fernandez, A., Casals, M.R., Colubi, A., Corral, N., Garcıa-Barzana, M., Gil, M.A., Gonzalez-Rodrıguez, G., Lopez, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., de la Rosa de Saa, S., Sinova, B.: Rejoinder on “A distance-based statistical analysis of fuzzy number-valued data”. *Int. J. Approx. Reas.* 55, 1601–1605 (2014)
3. Colubi, A., Domnguez-Menchero, J.S., Lpez-Daz, M., Ralescu, D.A.: On the formalization of fuzzy random variables. *Inf. Sci.* 133, 3–6 (2001)
4. Colubi, A., Lopez-Dıaz, M., Domınguez-Menchero, J.S. and Gil, M.A.: A generalized Strong Law of Large Numbers. *Probab. Theory Relat. Fields* 114, 401–417 (1999)
5. De Campos, L., Gonzalez, A. A subjective approach for ranking fuzzy numbers. *Fuzzy Set Syst.* 29, 145–153 (1989)
6. De la Rosa de Saa, S., Gil, M.A., Gonzalez-Rodrıguez, G., Lopez, M.T., Lubiano M.A.: Fuzzy rating scale-based questionnaires and their statistical analysis. *IEEE Trans. Fuzzy Syst.* 23(1), 111–126 (2015)
7. Diamond, P., Kloeden, P. Metric spaces of fuzzy sets. *Fuzzy Sets Syst.* 35, 241–249 (1990)
8. Donoho, D.L.: Breakdown properties of multivariate location estimators. Pd.D. quantifying paper, Dept. Statistics, Harvard University (1982)
9. Donoho, D.L., Huber, P.J.: The notion of breakdown point. In P.J. Bickel, K. Doksum, J.L. Hodges Jr. (Eds.), *A Festschrift for Erich L. Lehmann*, Wadsworth, Belmont. 157–184 (1983)
10. Frechet, M.: Les elements aleatoires de nature quelconque dans un espace distancie. *Ann. L’Inst. H. Poincare* 10, 215–310 (1948)
11. Goetschel Jr, R., Voxman, W. Elementary fuzzy calculus. *Fuzzy Sets Syst.* 18, 31–43 (1986)
12. Gonzalez-Rodrıguez, G., Colubi, A. and Gil, M.A.: Fuzzy data treated as functional data. A one-way ANOVA test approach. *Comp. Stat. Data Anal.* 56, 943–955 (2012)
13. Hampel, F.R.: Contributions to the Theory of Robust Estimation. Ph.D. Thesis, University of California, Berkeley (1968)

14. Körner, R.: On the variance of fuzzy random variables. *Fuzzy Sets Syst.* 92, 83–93 (1997)
15. Körner, R., Näther, W.: On the variance of random fuzzy variables. In Bertoluzza, C. *et al* (Eds.) *Statistical Modeling, Analysis and Management of Fuzzy Data*, 22–39. Physica-Verlag, Heidelberg (2002)
16. López-Díaz, M., Gil, M.A.: Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. *J. Stat. Plan. Infer.* 74, 11–29 (1998)
17. Lubiano, M.A., Gil, M.A.: Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Statistical Papers* 40(3), 277–295 (1999)
18. Lubiano, M.A., Gil, M.A., López-Díaz, M., López, M.T.: The  $\lambda$ -mean squared dispersion associated with a fuzzy random variable. *Fuzzy Sets Syst.* 111, 307–317 (2000)
19. Maronna, R.A., Martin, R.D., Yohai, V.J. *Robust Statistics: Theory and Methods*. Wiley (2006)
20. Ming, M. On embedding problems of fuzzy number spaces: Part 4. *Fuzzy Sets Syst.* 58, 185–193 (1993)
21. Nasibov, E.N. To linear transformations with fuzzy arguments. *Trans. Acad. Sciences Azerbaijan, Ser. Phys-tecn. and Math. Sciences.* 6, 164–169 (1989)
22. Puri, M.L., Ralescu, D.A.: The concept of normality for fuzzy random variables. *Ann. Probab.* 11, 1373–1379 (1985)
23. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. *J. Math. Anal. Appl.* 114, 409–422 (1986)
24. Ramos-Guajardo, A.B., Lubiano, M.A.: K-sample tests for equality of variances of random fuzzy sets. *Compu. Stat. Data Anal.* 56(4), 956–966 (2012)
25. Rousseeuw, P.J., Croux, C.: Alternatives to the Median Absolute Deviation. *J. Am. Statist. Assoc.* 88(424), 1273–1283 (1993)
26. Sinova, B., Gil, M.A., Colubi, A., Van Aelst, S.: The median of a random fuzzy number. The 1-norm distance approach. *Fuzzy Sets Syst.* 200, 99–115 (2012)
27. Sinova, B., De la Rosa de Súa, S., Gil, M.A.: A generalized  $L^1$ -type metric between fuzzy numbers for an approach to central tendency of fuzzy data. *Inform. Sci.* 242, 22–34 (2013)
28. Vitale, R.A.  $L_p$  metrics for compact, convex sets. *J. Approx. Theor.* 45, 280–287 (1985)
29. Yager, R.R. A procedure for ordering fuzzy subsets of the unit interval. *Inform. Sciences*, 24 143–161 (1981)
30. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part 1. *Inform. Sci.* 8, 199–249 (1975); Part 2. *Inform. Sci.* 8, 301–353 (1975); Part 3. *Inform. Sci.* 9, 43–80 (1975)