Location-free robust scale estimates for fuzzy data

Sara de la Rosa de Sáa, María Asunción Lubiano, Beatriz Sinova, Peter Filzmoser and María Ángeles Gil

***Abstract*—In analyzing fuzzy-valued imprecise data statistically, scale measures/estimates play an important role. Scale measures/estimates of datasets are often considered, among others, to descriptively summarize them, to compare the dispersion or the spread of different datasets, to standardize data, to state rules for detecting outliers, to formulate regression objective functions, and so on.**

To be robust, an estimate of scale should have a finite breakdown point close to 50% (i.e., around half data should be replaced by 'outliers' to make the estimate break down, either in the sense of exploding to infinity or imploding to zero). In this respect, the Median Distance Deviation about the median (MDD) for fuzzy datasets has already been introduced and its robust behaviour has been proved.

In contrast to the real-valued case, computation of the MDD for fuzzy data is much more complex and cannot be exactly but approximately performed in general. These computational inconveniencies are mainly associated with the fact that, in general, the 'median of the fuzzy dataset' cannot be exactly calculated, but simply approximated through some levels, and it does not preserve the shape of the fuzzy data. The same happens with the distances between data and the approximate median. Consequently, the use of location-free scale measures would be especially appropriate-to-use in this fuzzy-valued environment.

This paper aims to extend some robust global scale estimates, and to prove that the extension remains robust. Furthermore, it will be shown that these estimates can be easily and exactly computed for fuzzy trapezoidal data, the assumption of considering trapezoidal data not implying an important loss of generality in the setting of scale estimation.

***Index Terms*—finite sample breakdown point, fuzzy number-valued data, distance between fuzzy data, random fuzzy numbers, robust scale estimate, scale estimates of fuzzy data**

I. Introduction

In everyday life, there exist many data related to opinions, quality ratings, valuations, perceptions, etc. which cannot be appropriately expressed by using real numbers. Actually, most of these data are intrinsically imprecise and they can often be properly described and modeled by means of fuzzy numbers. The space of fuzzy numbers means a very rich scale, which is doubly (vertically and horizontally) continuous.

This research has been partially supported by the Principality of Asturias/FEDER Grant GRUPIN-IDI2018-000132 and the Spanish Ministry of Economy and Competitiveness Grant MTM2015-63971-P. Their support is gratefully acknowledged.

De la Rosa de Sáa is with the Oficina de Evaluación de Tecnologías Sanitarias del Principado de Asturias (OETSPA) 33001 Oviedo, Spain, Lubiano, Sinova and Gil are with the Department of Statistics, O.R. and D.M. of the University of Oviedo, 33007 Oviedo, Asturias, Spain, e-mail: magil@uniovi.es, (see http://bellman.ciencias.uniovi.es/SMIRE), and Filzmoser is with the Institut für Stochastik und Wirtschaftsmathematik, Technische Universität Wien, Wiedner Hauptstraße 8-10/E105, 1040 Wien, Austria.

Manuscript received ; revised ; accepted

On the other hand, random fuzzy numbers (RFNs for short, see Puri and Ralescu [1] for the seminal reference about) constitute a well-stated model for the random mechanism generating fuzzy number-valued data within a probabilistic setting. They integrate both randomness and fuzziness, so the random generation of experimental data and the fuzzy-valued imprecise nature of these data are assured.

In condensing the relevant information about the distribution of an RFN, one usually pays attention to two key aspects: the measurement of the central tendency or location and the representative measurement of its dispersion or its spread.

To summarize the location of an RFN, some measures have already been suggested. The so called Aumann-type mean (see Puri and Ralescu [1]) is an extended mean value that preserves all the main valuable properties of the mean of a random variable and it is coherent with the usual arithmetic with fuzzy numbers. Nevertheless, it also preserves an extremely sensitive behaviour to changes and the presence of atypical values (outliers). In the last years, extensions of robust location measures have been introduced and examined, namely, the 1-norm and the wabl/ldev/rdev medians (see Sinova *et al.* [2], [3], [4]), and the location M-estimates (see Sinova *et al.* [5]).

To summarize the representative measurement of the dispersion or the spread for fuzzy datasets, the best known measure is the Fréchet-type variance (see Körner [6] and Lubiano *et al.* [7]), and the associated standard deviation. As for the Aumann-type mean, the Fréchet-type variance preserves all the main valuable properties of the variance of a random variable, but it also preserves an extremely sensitive behaviour to data changes or the presence of outliers. Recently, a robust scale measure has been introduced (see De la Rosa de Sáa *et al.* [8]): the Median Distance Deviation (MDD for short) about the fuzzy median. Its robust behaviour has been shown in terms of its finite sample breakdown point and a kind of real-valued extension of the sensitivity curves. However, in contrast to the real-valued case, the computation of the MDD involves some complications. This is generally due to the fact that:

- the exact computation of the median is rather unfeasible; it requires the levelwise computation of some real-valued medians, so that in practice the median should be necessarily approximated through a finite number of its level sets;
- as a consequence, the exact computation of the distance between each fuzzy data and the median is rather unfeasible; actually, unlike what happens with M-estimates of location for fuzzy data, the median does not usually preserve the shape of fuzzy data and distances involved

in the MDD require levelwise computations and approximations.

To avoid the last drawbacks, this paper aims to present some estimates of scale for fuzzy data that are location-free and, apart from preserving some of the most important features of the MDD in [8], they also achieve the maximum finite sample breakdown point and have bounded real-valued extended sensitivity curves. Moreover, they can usually be exactly computed and have a simple and explicit formula (which as for the real-valued case entails a reduction in the computation time consumption).

This paper is organised so that in Section II the arithmetic with fuzzy numbers and the concept of RFN are recalled. Two metrics between fuzzy numbers to be used in this article, as well as two location measures (the Aumann-type mean, see Puri and Ralescu [1], and the 1-norm median, see Sinova *et al.* [2]) and two location-based scale measures (the Fréchet-type variance, see Körner [6] and Lubiano *et al.* [7], and the MDD about the 1-norm median, see De la Rosa de Sáa *et al.* [8]) are also recalled. Section III presents the three location-free estimates of scale proposed as alternatives to the MDD, illustrates them by means of a real-life example, and analyzes some of their properties as scale measures. The robust behavior of all these estimates is formally proved in Section IV by means of the exact and general computation of their finite sample breakdown point. Finally, on the basis of realistically inspired simulation developments, in Section V an approximation of the sample breakdown point is given and the real-valued extended sensitivity curves are also approximated for simulated 'symmetric' and 'asymmetric' distributions. The paper ends with a few comments on future directions.

II. Preliminary tools

As we have already commented, imprecise data can often be properly expressed and modeled by means of fuzzy numbers. In addition to the imprecision associated with interval values, fuzzy values allow us to capture gradualness (i.e., they can distinguish different degrees of compatibility of the real values in each interval with the imprecise valuation). In this section, the concept of fuzzy number and the usual fuzzy arithmetic based on Zadeh's extension principle [9] are recalled. In Section II-B we can find two metrics between fuzzy numbers which will be used in this paper. A formal model for the random mechanism generating fuzzy number-valued data along with some location measures and location-based scale measures for fuzzy datasets are recalled in Section II-C.

A. The space of fuzzy numbers

Fuzzy numbers (in the literature also referred to as fuzzy intervals) are formalized as follows:

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

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

with 'cl' *denoting the topological closure, is a nonempty compact interval. $\widetilde{U}_1$ is referred to as the* ***core*** *of $\widetilde{U}$ and $\{x \in \mathbb{R} : \widetilde{U}(x) > 0\}$ is called the* ***support*** *of $\widetilde{U}$.*

For each $x \in \mathbb{R}$, the value $\widetilde{U}(x)$ can be interpreted as the 'degree of compatibility of x with the property or valuation associated with $\widetilde{U}$'.

A useful example of fuzzy numbers are the *trapezoidal* ones, $\text{Tra}(a,b,c,d)$ with $a \leq b \leq c \leq d$, where $\big(\text{Tra}(a,b,c,d)\big)_\alpha = [\alpha \cdot b + (1-\alpha)\cdot a, \alpha \cdot c + (1-\alpha)\cdot d]$. Trapezoidal fuzzy numbers are very frequently used in practice because of their ease of drawing, interpreting and computing.

The space of (bounded) fuzzy numbers will be denoted by $\mathcal{F}_c^*(\mathbb{R})$. Real numbers and (nonempty) compact intervals can be viewed as special trapezoidal fuzzy numbers with $a = b = c = d$ and $a = b$, $c = d$, respectively.

The statistical analysis of fuzzy data demands the use of a suitable arithmetic to handle them. Namely, the two elementary operations required in performing statistics with fuzzy data are the sum between fuzzy numbers and the product of fuzzy numbers by scalars. The most usual and natural fuzzy arithmetic extends the usual arithmetic with real numbers and intervals, that is, the arithmetic based on Zadeh's extension principle [9]. Therefore, the sum and the product with fuzzy data are formalized as the levelwise extensions of the usual operations between intervals. More concretely, given $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ and $\gamma \in \mathbb{R}$,

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

$$(\widetilde{U} + \widetilde{V})_\alpha = \big[\inf \widetilde{U}_\alpha + \inf \widetilde{V}_\alpha, \sup \widetilde{U}_\alpha + \sup \widetilde{V}_\alpha\big],$$

that is, the Minkowski sum of $\widetilde{U}_\alpha$ and $\widetilde{V}_\alpha$.

Definition II.3. *The* ***product*** *of $\widetilde{U}$* ***by the scalar*** *γ is defined as the fuzzy number $\gamma \cdot \widetilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ such that for each $\alpha \in [0,1]$*

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

When the space of fuzzy numbers $\mathcal{F}_c^*(\mathbb{R})$ is endowed with the two preceding operations, it does not have a linear but a semilinear (actually a conical) structure, in contrast to what happens in the real-valued case. Consequently, there is no definition for the difference between fuzzy numbers which is simultaneously well-defined and preserves the fact that $(\widetilde{U} - \widetilde{V}) + \widetilde{V} = \widetilde{U}$ whatever $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ may be. Nevertheless, we can make use of appropriate metrics between fuzzy numbers, and this will allow us to consider the scale estimates extension in this paper.

B. Some metrics between fuzzy numbers

Throughout this paper, we will consider two extensions of the Euclidean distance in $\mathbb{R}$.

The metric which is to be mostly employed in this paper is the L^1-type 1-norm distance (see Diamond and Kloeden [10]) which is given by

Definition II.4. (Diamond and Kloeden [10]) *Given $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$, the **1-norm distance** between $\widetilde{U}$ and $\widetilde{V}$ is given by*

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

The computation of the ρ_1-distance between two fuzzy numbers can usually be exactly performed, although both the computation and the resulting expression can frequently be rather cumbersome, but in case of trapezoidal fuzzy numbers as it will be later shown.

The L^2-type 2-norm distance (see Diamond and Kloeden [10]) is the most commonly used to formalize the standard deviation, and it is defined as follows:

Definition II.5. (Diamond and Kloeden [10]) *Given $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c^*(\mathbb{R})$, the **2-norm distance** between $\widetilde{U}$ and $\widetilde{V}$ is given by*

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

The ρ_2-distance between two of the most commonly considered types of fuzzy numbers can usually be exactly computed, and both the computation and the resulting expression are not very complex (see, for instance, Lubiano *et al.* [11]).

C. Random fuzzy numbers

When fuzzy data associated with random experiments are considered and a statistical analysis of these data is going to be carried out, a well-supported mathematical model for the random mechanisms generating these data is needed. Random fuzzy numbers have been shown to be a sound model for this purpose.

Consider a random experiment which is mathematically modeled by means of a probability space $(\Omega, \mathcal{A}, P)$.

Definition II.6. (Puri and Ralescu [1]) *A **random fuzzy number** (for short RFN) associated with $(\Omega, \mathcal{A}, P)$ is a mapping $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R})$ such that for all $\alpha \in [0,1]$ the α-level mapping $\mathcal{X}_\alpha : \Omega \to \mathcal{P}(\mathbb{R})$ (power set of the space of real numbers) given by $\mathcal{X}_\alpha(\omega) = \big(\mathcal{X}(\omega)\big)_\alpha$ is a compact random interval, that is, the real-valued mappings* $\inf \mathcal{X}_\alpha$ *and* $\sup \mathcal{X}_\alpha$ *are random variables.*

Remark II.1. Random fuzzy numbers have been introduced in a more general dimension by Puri and Ralescu [1], who coined them as fuzzy random variables. These random elements have also been referred to in the literature as random fuzzy sets (see, for instance, [12], [13], [14]) and random upper semicontinuous functions (see, for instance, [15], [16], [17]).

Remark II.2. It is known that a mapping $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R})$ is an RFN if and only if it is a Borel-measurable mapping w.r.t. the Borel σ-field generated on $\mathcal{F}_c^*(\mathbb{R})$ by the topology induced by several different metrics, like those in Definitions II.4 and II.5. This Borel-measurability allows us to refer to the *distribution induced by an RFN*, the *stochastic independence of RFNs*, and so on, without needing to state these notions expressly.

In summarizing the (induced) distribution of an RFN, the best known central tendency measure is the Aumann-type mean (see Puri and Ralescu [1]), which fulfills many valuable properties and is supported by Strong Laws of Large Numbers.

Consider the RFN $\mathcal{X}$ and a sample of individuals $(\omega_1, \dots, \omega_n)$, and let $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \dots, \widetilde{x}_n)$ be the associated sample of fuzzy data, that is, $\widetilde{x}_i = \mathcal{X}(\omega_i)$ for $i = 1, \dots, n$.

Definition II.7. (Puri and Ralescu [1]) *The **(sample) Aumann-type mean** of the fuzzy dataset $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \dots, \widetilde{x}_n)$ is defined as the fuzzy number $\overline{\widetilde{\mathbf{x}}}_n$ such that for all $\alpha \in [0,1]$*

$$\left(\overline{\widetilde{\mathbf{x}}}_n\right)_\alpha = \left[\frac{1}{n} \sum_{i=1}^n \inf(\widetilde{x}_i)_\alpha, \frac{1}{n} \sum_{i=1}^n \sup(\widetilde{x}_i)_\alpha \right].$$

With the aim of avoiding the strong influence of atypical values or changes on the Aumann-type mean of an RFN, the following robust location measure was introduced and examined a few years ago.

Definition II.8. (Sinova *et al.* [2]) *The **(sample) 1-norm median** of the fuzzy dataset $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \dots, \widetilde{x}_n)$ is defined as the fuzzy number $\widehat{\mathrm{Me}}(\widetilde{\mathbf{x}}_n)$ such that for all $\alpha \in [0,1]$*

$$\left(\widehat{\mathrm{Me}}(\widetilde{\mathbf{x}}_n)\right)_\alpha = \Big[\mathrm{Me}\{\inf(\widetilde{x}_1)_\alpha, \dots, \inf(\widetilde{x}_n)_\alpha\}, \mathrm{Me}\{\sup(\widetilde{x}_1)_\alpha, \dots, \sup(\widetilde{x}_n)_\alpha\} \Big],$$

with $\mathrm{Me}\{\cdot\}$ *denoting the sample median of the corresponding real-valued dataset, and by following the most usual convention of choosing the mid-point of the interval of medians in case it is not unique.*

The 1-norm median of an RFN preserves all the most remarkable properties of the median of a random variable, including its high robustness w.r.t. either changes in data or the presence of outliers.

The two preceding location estimates for fuzzy data are both equivariant by scale and location. Moreover, whereas the sample Aumann-type mean is the fuzzy number minimizing the mean squared ρ_2-distance between sample fuzzy data and a fuzzy number, the sample 1-norm median is (one of) the fuzzy number(s) minimizing the mean ρ_1-distance between sample fuzzy data and a fuzzy number.

In addition to the central tendency, the distribution of an RFN is often summarized by means of scale measures. Following the ideas by Bickel and Lehmann in the real-valued case for dispersion and spread measures [19], [20], by a ***scale measure/estimate*** we mean a nonnegative measure/estimate which is (nonnegative) scale equivariant, and (fuzzy) location and sign invariant (often referred to as affine equivariant), that is, a measure $\tau(\mathcal{X})$ such that

$$\tau(c \cdot \mathcal{X}) = |c| \cdot \tau(\mathcal{X}) \quad \text{and} \quad \tau(\mathcal{X} + \widetilde{U}) = \tau(\mathcal{X})$$

whatever $c \in \mathbb{R}$ and $\widetilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ may be. Furthermore, it is commonly assumed that in case the distribution of $\mathcal{X}$ is degenerate at a fuzzy number, then $\tau(\mathcal{X}) = 0$.

Körner [6] and Lubiano *et al.* [7] have introduced the Fréchet-type variance, and the associated standard deviation can be trivially derived.

Definition II.9. (De la Rosa de Sáa *et al.* [13], based on Körner [6] and Lubiano *et al.* [7]) *The* ***(sample) Fréchet-type Standard Deviation*** *estimator,* $\widehat{\rho_2\text{-SD}}_n$*, associates with the sample of fuzzy data* $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$ *the real number given by*

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

The standard deviation defined above preserves all the most remarkable properties of that of a random variable, including its high sensitivity w.r.t. either changes in data or the presence of outliers. Aiming to reduce such a sensitivity as much as possible, De la Rosa de Sáa *et al.* [8] have introduced the following scale measure.

Definition II.10. (De la Rosa de Sáa *et al.* [8]) *The* ***(sample) Median Distance Deviation about the 1-norm median estimator****,* $\widehat{\rho_1\text{-MDD}}_n$*, associates with the sample of fuzzy data* $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$ *the real number given by*

$$\widehat{\rho_1\text{-MDD}}(\widetilde{\mathbf{x}}_n) = \mathrm{Me}\left\{\rho_1\left(\widetilde{x}_1, \widehat{\widetilde{\mathrm{Me}}}(\widetilde{\mathbf{x}}_n)\right), \ldots, \rho_1\left(\widetilde{x}_n, \widehat{\widetilde{\mathrm{Me}}}(\widetilde{\mathbf{x}}_n)\right)\right\},$$

where in case the real-valued Me *is not unique the usual convention of choosing the mid-point of the interval of medians will be employed.*

The Median Distance Deviation about the 1-norm median preserves all the essential properties of the MDD for random variables. De la Rosa de Sáa *et al.* [13] have proved such properties along with the fact that, whereas the number of sample data which should be replaced by 'outliers' to make the estimate break down (either in the sense of exploding to infinity or imploding to zero) equals 1 for SD, it equals $\lfloor n/2 \rfloor$ for MDD (with $\lfloor \cdot \rfloor$ denoting the floor function), which corresponds to the highest possible value for the finite sample breakdown point.

III. Some location-free scale estimates as alternatives to the MDD for fuzzy data

Since fuzzy numbers represent imprecise values, in general one cannot rank them according to their 'magnitude'. Total orderings can be stated in $\mathcal{F}_c^*(\mathbb{R})$ (see, for instance, [21], [22] for suggested orderings), but none can be universally accepted, whence the idea of right and left deviations cannot make rigourous general sense in the fuzzy numbers setting. Nevertheless, as for the case of real-valued data, one can assert that the robust MDD takes a kind of 'symmetric' view of dispersion because of involving distances between data and a central value of the dataset. Therefore, the MDD is not a natural scale measure in dealing with 'asymmetric' distributions.

Furthermore, in the case of fuzzy datasets the computation of the MDD involves some added concerns, namely: the one associated with the computation of the fuzzy-valued 1-norm median, which should be approximated through a finite number of levels and does not usually preserve data shape, and, consequently, the one associated with the computation of distances between fuzzy data and the 1-norm median, which should also be approximated on the basis of the distances for a finite number of levels.

The aim of this paper is to introduce some scale estimates for fuzzy data with maximum finite sample breakdown point, that is, $\lfloor n/2 \rfloor / n$ (the one for MDD), properly behaving with asymmetric distributions and involving lower computational cost and complexity. For this purpose, the extension of three scale estimates for real-valued datasets suggested by Rousseeuw and Croux [23], [24] is now to be considered.

A. Extended location-free scale estimates for fuzzy data

Let $\mathcal{X}$ be an RFN and let $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$ be a sample of observations from it.

Definition III.1. *The* ***(sample) scale estimator*** $\widehat{\rho_1\text{-S}}_n$ *associates with* $\widetilde{\mathbf{x}}_n$ *the real number given by*

$$\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n) = \underline{\mathrm{Me}}_i\left\{\overline{\mathrm{Me}}_j\left\{\rho_1(\widetilde{x}_i, \widetilde{x}_j)\right\}\right\},$$

where $\underline{\mathrm{Me}}$ *is a low median, that is, the* l_n*-th order statistic with* $l_n := \lfloor (n+1)/2 \rfloor$ *and* $\overline{\mathrm{Me}}$ *is a high median, that is, is the* h_n*-th order statistic with* $h_n := \lceil (n+1)/2 \rceil = \lfloor n/2 \rfloor + 1$*, where* $\lfloor \cdot \rfloor$ *and* $\lceil \cdot \rceil$ *denote the floor and ceiling functions, respectively.*

Therefore, for each i *we compute the high median of* $\{\rho_1(\widetilde{x}_i, \widetilde{x}_j) : j = 1, \ldots, n\}$*. This leads to* n *real numbers, the low median of which gives the estimate* $\widehat{\rho_1\text{-S}}_n$*.*

Definition III.2. *The* ***(sample) scale estimator*** $\widehat{\rho_1\text{-Q}}_n$ *associates with* $\widetilde{\mathbf{x}}_n$ *the real number given by*

$$\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n) = \{\rho_1(\widetilde{x}_i, \widetilde{x}_j) : i < j\}_{(m_n)},$$

where $m_n := \binom{h_n}{2}$*.*

That is, it corresponds to the m_n*-th order statistic of the* $\binom{n}{2}$ *distances between each two different fuzzy data.*

Definition III.3. *The* ***(sample) scale estimator*** $\widehat{\rho_1\text{-T}}_n$ *associates with* $\widetilde{\mathbf{x}}_n$ *the real number given by*

$$\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n) = \frac{1}{h_n}\sum_{r=1}^{h_n}\left\{\overline{\mathrm{Me}}_j\left\{\rho_1(\widetilde{x}_i, \widetilde{x}_j)\right\}\,;\; i = 1, \ldots, n\right\}_{(r)}.$$

For each i *the estimate calculates the high median of* $\{\rho_1(\widetilde{x}_i, \widetilde{x}_j) : j = 1, \ldots, n\}$*, leading to* n *medians. Then, it computes the average of the* h_n *first ordered medians.*

Note that the above defined scale estimates have an explicit formula. Moreover, whereas $\widehat{\rho_1\text{-MDD}}_n$ and $\widehat{\rho_2\text{-SD}}_n$ involve the previous computation of a location estimate (the sample 1-norm median and the sample Aumann type mean, respectively), those in Definitions III.1, III.2 and III.3 do not require any reference to the center of the distribution since they only take into account distances between observations.

To illustrate the preceding scale estimates we now compute them in a real-life example.

B. Illustrative real-life example

Consider the sample of 23 trapezoidal fuzzy numbers collected in Table I and displayed on the left of Figure 1.

TABLE I
SAMPLE OF 23 TRAPEZOIDAL FUZZY NUMBERS WITH ONE (BOLD) OUTLIER, AND $(a_i, b_i, c_i, d_i) \equiv \mathrm{Tra}(a_i, b_i, c_i, d_i) = \widetilde{x}_i$

(a_i, b_i, c_i, d_i)	(a_i, b_i, c_i, d_i)	(a_i, b_i, c_i, d_i)
(8.4,9,10,10)	(8,8.5,8.5,9)	(8,8.5,9.2,9.2)
(3,3,3.45,4)	(6,6,6.6,7.7)	(7,8,9,9)
(8.6,10,10,10)	(8.1,8.2,8.6,9)	(9,10,10,10)
(10,10,10,10)	(9.6,9.8,10,10)	(8,10,10,10)
(8.9,9.4,10,10)	(9.2,9.8,10,10)	(6,7,9,10)
(10,10,10,10)	(5.2,5.4,5.65,6)	(8.6,9.15,9.75,10)
(8.7,9.4,10,10)	(8,9,10,10)	(5.1,6,6.75,7.3)
(9,10,10,10)	(10,10,10,10)	

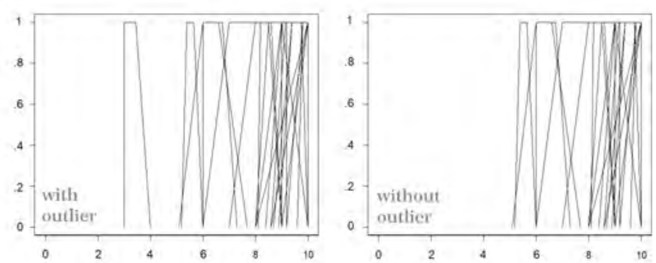


Fig. 1. Sample of trapezoidal fuzzy numbers with (on the left) and without (on the right) outlier $\widetilde{x}_2$

They correspond to the fuzzy responses provided by 23 Grade 4 students (nine to ten years old) to the question in the well-known TIMSS-PIRLS 2011 Student's Questionnaire "How much do you agree with this statement about learning mathematics: $M.2$. My teacher is easy to understand?". In the original version of the Questionnaire, items are designed so that responses are to be given by considering a 4-point Likert scale (labels of the potential responses being DISAGREE A LOT, DISAGREE A LITTLE, AGREE A LITTLE, and AGREE A LOT). Nevertheless, several items have been adapted to produce more expressive and informative responses and statistical conclusions by considering a fuzzy rating scale in Hesketh *et al.*'s sense [25] (see Figure 2 for the adapted version associated with Item $M.2$ for the paper-and-pencil form). The adapted questionnaire was conducted in 2014 on Grade 4 students of Colegio San Ignacio, in Oviedo, and the dataset in Table I has been supplied by the students filling out the paper-and-pencil form (some others have filled out a computerized form).

The bold datum in the table, $\widetilde{x}_2 = \mathrm{Tra}(3, 3, 3.45, 4)$, clearly stands out from the rest of the values and can be viewed as an 'outlier'.

Table II gathers the values of the scale estimates recalled in Section II and those defined in this section for the whole dataset (with the outlier) and also for the reduced dataset in which the outlier has been removed. We can observe the high sensitivity of $\widehat{\rho_2\text{-SD}}(\widetilde{\mathbf{x}}_n)$ to the presence of this outlier in contrast to the more robust behavior of $\widehat{\rho_1\text{-MDD}}(\widetilde{\mathbf{x}}_n)$ and the highly robust behavior of $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n)$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n)$ and $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n)$, since the last three values do not change when the outlier is removed from the dataset.

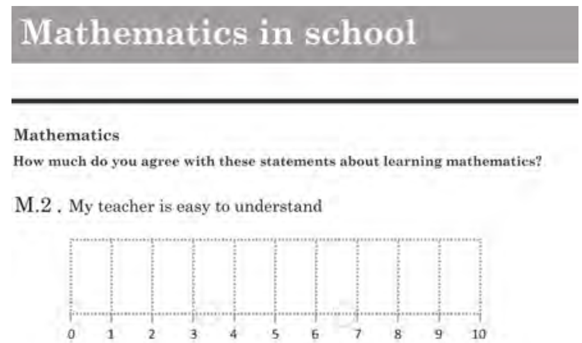


Fig. 2. Example of an item in the paper-and-pencil form of the adapted questionnaire

TABLE II
THE EFFECT OF REMOVING THE OUTLIER IN SOME SCALE ESTIMATES

Scale measure	Dataset WITH outlier	Dataset WITHOUT outlier
$\widehat{\rho_2\text{-SD}}(\widetilde{\mathbf{x}}_n)$	1.69	1.29
$\widehat{\rho_1\text{-MDD}}(\widetilde{\mathbf{x}}_n)$	0.60	0.48
$\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n)$	0.63	0.63
$\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n)$	0.38	0.38
$\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n)$	0.49	0.49

C. The alternative estimates are scale estimates

It can be proved that the estimates introduced in this section are in fact scale estimates as above indicated. Thus,

Proposition III.1. *$\widehat{\rho_1\text{-S}_n}$, $\widehat{\rho_1\text{-Q}_n}$ and $\widehat{\rho_1\text{-T}_n}$ are nonnegative and affine equivariant estimators, and in case the observations in the sample are equal, that is, $\widetilde{\mathbf{x}}_n = (\widetilde{x}, \overset{(n\,\text{times})}{\ldots}, \widetilde{x})$, then $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n) = 0$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n) = 0$ and $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n) = 0$.*

Remark III.1. As for the real-valued case, and for the MDD estimator, either $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n)$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n)$ or $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n)$ vanishing does not generally entail that all observations in the sample $\widetilde{\mathbf{x}}_n$ coincide. To illustrate this assertion consider, for instance, the fuzzy-valued sample $\widetilde{\mathbf{x}}_5 = (\mathrm{Tri}(0,1,2), \mathrm{Tri}(1,2,3), \mathrm{Tri}(1,2,3), \mathrm{Tri}(1,2,3), \mathrm{Tri}(2,3,4))$ with $\mathrm{Tri}(a,b,c) = \mathrm{Tra}(a,b,b,c)$. It is straightforward to check that the values of the scale estimates in Definitions III.1, III.2 and III.3 equal zero. Consequently, the vanishing of scale estimates in Definitions III.1, III.2 and III.3 does not necessarily ensure the lack of variability. This is due to the fact that these estimates should be viewed as representative measures of the spread of sample fuzzy data (as understood by Bickel and Lehmann [20]), instead of measures of variability.

D. Remarks on the use of trapezoidal fuzzy data and real-life example-based remark on the influence of the shape of fuzzy data

As it has already been commented, trapezoidal fuzzy numbers are easy to draw, interpret and compute. In the last respect, all the computations involved in the location-free estimates in this section can be straightforwardly carried out for trapezoidal fuzzy numbers.

In this way, the ρ_1- and the ρ_2-distances between two trapezoidal fuzzy numbers $\mathrm{Tra}(a_1, b_1, c_1, d_1)$ and

$\mathrm{Tra}(a_2, b_2, c_2, d_2)$ can be exactly obtained and they equal

$$\rho_1\big(\mathrm{Tra}(a_1, b_1, c_1, d_1), \mathrm{Tra}(a_2, b_2, c_2, d_2)\big)$$
$$= G_{\rho_1}(a_1 - a_2, b_1 - b_2) + G_{\rho_1}(c_1 - c_2, d_1 - d_2),$$

where

$$G_{\rho_1}(x, y) = \begin{cases} \dfrac{x|x| - y|y|}{4(x-y)} & \text{if } x \neq y \\ \dfrac{|y|}{2} & \text{otherwise.} \end{cases},$$

$$\rho_2\big(\mathrm{Tra}(a_1, b_1, c_1, d_1), \mathrm{Tra}(a_2, b_2, c_2, d_2)\big)$$
$$= \sqrt{G_{\rho_2}(a_1 - a_2, b_1 - b_2) + G_{\rho_2}(c_1 - c_2, d_1 - d_2)},$$

where

$$G_{\rho_2}(x, y) = \frac{x^2 + y^2 + xy}{6}.$$

It should be also pointed out that in case all fuzzy data in the sample are trapezoidal, $\widetilde{x}_i = \mathrm{Tra}(a_i, b_i, c_i, d_i)$, then:

- the sample mean is also trapezoidal, more concretely,
$$\overline{\widetilde{\mathbf{x}}}_n = \mathrm{Tra}\left(\frac{1}{n}\sum_{i=1}^n a_i, \frac{1}{n}\sum_{i=1}^n b_i, \frac{1}{n}\sum_{i=1}^n c_i, \frac{1}{n}\sum_{i=1}^n d_i\right),$$
what substantially eases the exact computation of $\widehat{\rho_2\text{-SD}}_n$; the same happens for most of LR-fuzzy data (see [18]);
- however, contrary to what happens with the sample mean, in case all fuzzy data in the sample are trapezoidal one cannot generally anticipate the shape of the sample 1-norm median, which should be approximated through a large finite number of levels and does not usually share the trapezoidal shape. In consequence, the computation of $\widehat{\rho_1\text{-MDD}}_n$ becomes quite cumbersome and it is not exact; the same happens for most of LR-fuzzy data.

The last remark concerning the computation of $\widehat{\rho_1\text{-MDD}}_n$ strongly supports the convenience of introducing location-free scale estimates.

On the other hand, it should be pointed out that the assumption of fuzzy data having trapezoidal shape is not very restrictive in practice. This assertion has been inferentially proven for Fréchet's variance (see De la Rosa de Sáa *et al.* [26]) for the case study in Section III-B, and conclusions can be corroborated by means of different simulation studies (p-values of the test about the equality of variances are shown to be very high, so there is no significant difference between variances for different shapes).

And this can be also descriptively shown for the location-free scale estimates which have been introduced in this paper. More concretely, assume the fuzzy data from the illustrative example in Section III-B are joined to the fuzzy data for the remaining 45 students who completed the computerized form of the same questionnaire. Secondly, the linear 'arms' of their trapezoidal shape are replaced by other ones which are instances of the so-called LU-fuzzy numbers (see, for instance, Stefanini *et al.* [27]), so that the core and the support of the fuzzy data are preserved and they can represent close concepts or valuations (see Figure 3).

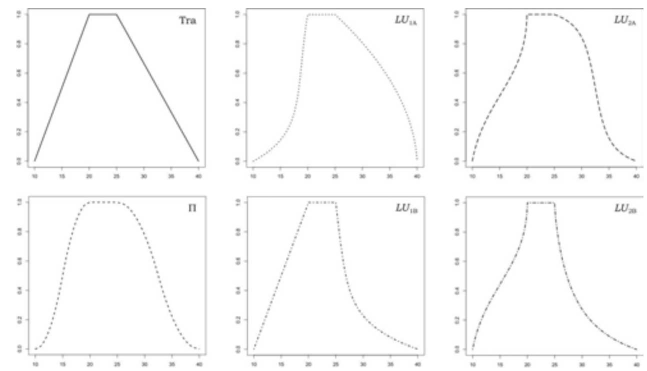


Fig. 3. Six types of fuzzy numbers sharing core $[20, 25]$ and support $(10, 40)$ and differing in shape. On the left, trapezoidal (top) and Π-curve (bottom), along with four different LU fuzzy numbers on the middle and the right

Then, the resulting datasets in http://bellman.ciencias.uniovi.es/SMIRE/FuzzyRatingScaleQuestionnaire-SanIgnacio.html lead to the outputs in Table III. Notice that values in each row scarcely differ. Similar conclusions would be drawn for other items and simulated examples, as shown in Section V-B.

TABLE III
SCALE ESTIMATES VALUES FOR THE RESPONSES TO ITEM $M.2$ IN SECTION III-B, DEPENDING ON THE CONSIDERED SHAPE

estimate \ shape	Tra	Π	LU_{1A}	LU_{1B}	LU_{2A}	LU_{2B}
$\widehat{\rho_2\text{-SD}}_n$	2.3419	2.3380	2.3012	2.3785	2.3357	2.3722
$\widehat{\rho_1\text{-MDD}}_n$	1.7374	1.7371	1.6887	1.7951	1.7332	1.7870
$\widehat{\rho_1\text{-S}}_n$	1.5000	1.5000	1.3991	1.5600	1.4867	1.5390
$\widehat{\rho_1\text{-Q}}_n$	0.8125	0.8063	0.8007	0.8250	0.8090	0.8269
$\widehat{\rho_1\text{-T}}_n$	1.1786	1.1785	1.1314	1.2317	1.1744	1.2253

Remarks concerning the computation of $\widehat{\rho_1\text{-MDD}}_n$ along with the small influence of the shape of fuzzy data on the scale estimates and the ease to design and fill out questionnaires allowing fuzzy responses, strongly reinforce the convenience of using trapezoidal fuzzy data and location-free scale estimates.

IV. FORMAL ANALYSIS OF THE ROBUSTNESS OF THE LOCATION-FREE SCALE ESTIMATES THROUGH THEIR FINITE SAMPLE BREAKDOWN POINTS

The situation in the example in Section III-B illustrates how strongly the presence of an atypical fuzzy datum influences the value of $\widehat{\rho_2\text{-SD}}(\widetilde{\mathbf{x}}_n)$, whereas such an influence is less relevant for $\widehat{\rho_1\text{-MDD}}_n$ and not relevant for $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n)$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n)$ or $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n)$.

In order to generally analyze whether or not this assertion is true, a popular and powerful tool is the breakdown point. Donoho and Huber [28] stated that "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" [29]. Although originally it was presented for location estimates, the concept has also been generalized to scale estimates.

A simple and intuitive definition of the breakdown point restricted to finite samples, the so-called *finite sample breakdown point* (fsbp for short), was introduced by Donoho [30] and Donoho and Huber [28]. For scale estimates the fsbp is defined as the minimum proportion of sample data which should be perturbed in order to let the estimate achieve either an arbitrarily large value or the value zero. The higher the breakdown point of an estimate, the more robust it is. Therefore, two situations are to be studied: the one consisting of contaminating the sample by means of outliers, which can make the estimate overestimate the true scale up to infinity (*explosion*), and the one consisting of contaminating the sample by means of inliers, which may result in underestimation of the true scale to zero (*implosion*). Notice that in dealing with location estimates only the explosion case makes sense.

Next, the replacement version of the finite sample breakdown point for scale estimates (see Donoho and Huber [28]) is adapted to deal with fuzzy data.

Definition IV.1. *For any sample of observations $\widetilde{\mathbf{x}}_n$ from an RFN $\mathcal{X}$, the* ***finite sample breakdown point of a scale estimate*** *$\widehat{\tau}(\widetilde{\mathbf{x}}_n)$ is defined by*

$$\mathrm{fsbp}(\widehat{\tau}(\widetilde{\mathbf{x}}_n)) = \min\left\{\mathrm{fsbp}^+(\widehat{\tau}(\widetilde{\mathbf{x}}_n)),\, \mathrm{fsbp}^-(\widehat{\tau}(\widetilde{\mathbf{x}}_n))\right\}$$

where

$$\mathrm{fsbp}^+(\widehat{\tau}(\widetilde{\mathbf{x}}_n)) = \min\left\{\frac{k}{n}; \sup_{\widetilde{\mathbf{y}}_{n,k}} \widehat{\tau}(\widetilde{\mathbf{y}}_{n,k}) = \infty\right\}$$

and

$$\mathrm{fsbp}^-(\widehat{\tau}(\widetilde{\mathbf{x}}_n)) = \min\left\{\frac{k}{n}; \inf_{\widetilde{\mathbf{y}}_{n,k}} \widehat{\tau}(\widetilde{\mathbf{y}}_{n,k}) = 0\right\}$$

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

The following theorems prove that if the considered sample of fuzzy observations $\widetilde{\mathbf{x}}_n$ does not contain two coinciding observations, then its fsbp equals $\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor$ for the three scale estimates introduced in Section III, which is the highest possible fsbp for a scale estimate, as it has already been said. Therefore, these estimators inherit the value of the fsbp from the real-valued case (see, for instance, Rousseeuw and Croux [24]).

Theorem IV.1. *For any sample of observations $\widetilde{\mathbf{x}}_n$ from an RFN $\mathcal{X}$ in which there are not two identical observations, we have that*

$$\mathrm{fsbp}^+(\widehat{\rho_1\text{-}\mathrm{S}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n+1}{2}\right\rfloor,\ \mathrm{fsbp}^-(\widehat{\rho_1\text{-}\mathrm{S}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor.$$

Therefore, the finite sample breakdown point of the scale estimate $\widehat{\rho_1\text{-}\mathrm{S}}(\widetilde{\mathbf{x}}_n)$ is given by

$$\mathrm{fsbp}(\widehat{\rho_1\text{-}\mathrm{S}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor,$$

which is the highest possible fsbp *of a scale estimate.*

Theorem IV.2. *For any sample of observations $\widetilde{\mathbf{x}}_n$ from an RFN $\mathcal{X}$ in which there are not two identical observations, we have that*

$$\mathrm{fsbp}^+(\widehat{\rho_1\text{-}\mathrm{Q}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n+1}{2}\right\rfloor,\ \mathrm{fsbp}^-(\widehat{\rho_1\text{-}\mathrm{Q}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor.$$

Therefore, the finite sample breakdown point of the scale estimate $\widehat{\rho_1\text{-}\mathrm{Q}}(\widetilde{\mathbf{x}}_n)$ is given by

$$\mathrm{fsbp}(\widehat{\rho_1\text{-}\mathrm{Q}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor,$$

which is the highest possible fsbp *of a scale estimate.*

Theorem IV.3. *For any sample of observations $\widetilde{\mathbf{x}}_n$ from an RFN $\mathcal{X}$ in which there are not two identical observations, we have that*

$$\mathrm{fsbp}^+(\widehat{\rho_1\text{-}\mathrm{T}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n+1}{2}\right\rfloor,\ \mathrm{fsbp}^-(\widehat{\rho_1\text{-}\mathrm{T}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor.$$

Therefore, the finite sample breakdown point of the scale estimate $\widehat{\rho_1\text{-}\mathrm{T}}(\widetilde{\mathbf{x}}_n)$ is given by

$$\mathrm{fsbp}(\widehat{\rho_1\text{-}\mathrm{T}}(\widetilde{\mathbf{x}}_n)) = \frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor,$$

which is the highest possible fsbp *of a scale estimate.*

Proofs of the preceding results can be found in the supplementary material.

V. Realistic simulation-based analyses of the robustness of the global scale estimates

A crucial thought at this stage is that there are not any suitable realistic models for the distribution of an RFN yet; this makes the simulation process a rather novel endeavor.

In the setting of fuzzy data, the notion of outlier for the real-valued case still makes sense: they are observations that are 'separated' from the majority of data because, among others, they have either a completely different 'location' or a completely different 'imprecision' (i.e., with a different scale on the core and the support, that is, 'wider' or 'narrower' than most of the data) or both.

Section V-A describes how to generate (non-contaminated) fuzzy data.

A. Simulating non-contaminated fuzzy data

The generation procedure of the (non-contaminated) sample is inspired by several real-life fuzzy datasets (see, for instance, De la Rosa de Sáa *et al.* [13], Sinova *et al.* [5]) involving the use of the so-called fuzzy rating scale (see Hesketh *et al.* [25]) and by applying goodness-of-fit techniques to model the distribution of some key real values of the core and support of the fuzzy data.

More concretely, to generate the non-contaminated fuzzy dataset from a random LU-valued fuzzy number $\mathcal{X} = LU(\inf\mathcal{X}_0, \inf\mathcal{X}_1, \sup\mathcal{X}_1, \sup\mathcal{X}_0)$ taking on fuzzy values characterized by 4-tuples, we follow the process in De la Rosa de Sáa *et al.* [13] and Lubiano *et al.* [11], [31].

To ease simulation and modelling, $\mathcal{X}$ is to be characterized by means of four real-valued random variables, X_1, X_2, X_3 and X_4, where

$$X_1 = \mathrm{mid}\,\mathcal{X}_1 = (\inf\mathcal{X}_1 + \sup\mathcal{X}_1)/2,$$
$$X_2 = \mathrm{spr}\,\mathcal{X}_1 = (\sup\mathcal{X}_1 - \inf\mathcal{X}_1)/2,$$
$$X_3 = \mathrm{lspr}\,\mathcal{X}_0 = \inf\mathcal{X}_1 - \inf\mathcal{X}_0,$$
$$X_4 = \mathrm{uspr}\,\mathcal{X}_0 = \sup\mathcal{X}_0 - \sup\mathcal{X}_1,$$

that is,

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

Fuzzy data have been generated so that

– $100 \cdot \omega_1\%$ of the data have been obtained by first considering a simulation from a simple random sample of size 4 from a beta $\beta(p, q)$ distribution, the ordered 4-tuple, (a, b, c, d), and finally computing the values of the x_i. The values from the beta distribution are re-scaled and translated to an interval $[0, 100]$.
– $100 \cdot \omega_2\%$ of the data have been obtained considering a simulation of four random variables $X_i = 100 \cdot Y_i$ as follows:

$$\begin{aligned} &Y_1 \sim \beta(p, q), \\ &Y_2 \sim \text{Uniform}\big[0, \min\{1/10, Y_1, 1 - Y_1\}\big], \\ &Y_3 \sim \text{Uniform}\big[0, \min\{1/5, Y_1 - Y_2\}\big], \\ &Y_4 \sim \text{Uniform}\big[0, \min\{1/5, 1 - Y_1 - Y_2\}\big]. \end{aligned}$$

– $100 \cdot \omega_3\%$ of the data have been obtained considering a simulation of four random variables $X_i = 100 \cdot Y_i$ as follows:

$$Y_1 \sim \beta(p, q),$$

$$Y_2 \sim \begin{cases} \text{Exp}(200) & \text{if } Y_1 \in [0.25, 0.75] \\ \text{Exp}(100 + 4\,Y_1) & \text{if } Y_1 < 0.25 \\ \text{Exp}(500 - 4\,Y_1) & \text{otherwise} \end{cases}$$

$$Y_3 \sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 - Y_2 \geq 0.25 \\ \gamma(4, 100 + 4\,Y_1) & \text{otherwise} \end{cases}$$

$$Y_4 \sim \begin{cases} \gamma(4, 100) & \text{if } Y_1 + Y_2 \geq 0.25 \\ \gamma(4, 500 - 4\,Y_1) & \text{otherwise.} \end{cases}$$

On the basis of simulated fuzzy data one corroborates in Section V-B what has been previously highlighted in Section III-D by means of a real-life example.

B. Simulation-based analysis of the influence of the shape of fuzzy data

As it has been commented in Section III-D, computations required to obtain the scales estimates $\widehat{\rho_1\text{-S}_n}$, $\widehat{\rho_1\text{-Q}_n}$ and $\widehat{\rho_1\text{-T}_n}$ are substantially easier if involved fuzzy data are trapezoidal. Moreover, assuming that available fuzzy data are trapezoidal does not entail a significant loss of generality in statistical summarization. This assertion has been corroborated in previous studies in connection with the Aumann-type mean and other central tendency measures (see Lubiano *et al.* [11], [32]) and the Fréchet-type variance (see De la Rosa de Sáa *et al.* [26]).

In this section, on the basis of the results from the simulation of fuzzy data which have been gathered in Table IV, one can immediately confirm that the shape of fuzzy data representing close concepts/valuations does not substantially affect the values of the scale estimates. In this respect, Table IV shows

- the values of the scale estimates $\widehat{\rho_2\text{-SD}_n}$, $\widehat{\rho_1\text{-MDD}_n}$, $\widehat{\rho_1\text{-S}_n}$, $\widehat{\rho_1\text{-Q}_n}$ and $\widehat{\rho_1\text{-T}_n}$ for a sample of n LU fuzzy data (with $LU \in \{\text{Tra}, \Pi, LU_{1A}, LU_{2A}, LU_{1B}, LU_{2B}\}$, $n = 20$ and $n = 100$) simulated by following the procedure described in Section V-A, where the considered weights are $\omega_1 = 0.8$, $\omega_2 = 0.1$ and $\omega_3 = 0.1$, and
- the mean squared errors over 1000 samples of size n.

Next section describes how to generate fuzzy outliers.

TABLE IV
SCALE ESTIMATES VALUES IN A SPECIFIC SAMPLE AND MEAN SQUARED ERROR OVER 1000 SAMPLES OF n SIMULATED FUZZY DATA WITH DIFFERENT SHAPES

EXAMPLE OF THE ESTIMATE FOR A SAMPLE
(sample size $n = 20$)

estimate \ shape	Tra	Π	LU_{1A}	LU_{1B}	LU_{2A}	LU_{2B}
$\widehat{\rho_2\text{-SD}_n}$	0.6471	0.6359	0.6392	0.6875	0.6457	0.6896
$\widehat{\rho_1\text{-MDD}_n}$	0.4390	0.4165	0.4282	0.4624	0.4376	0.4598
$\widehat{\rho_1\text{-S}_n}$	0.6786	0.6625	0.6944	0.7044	0.6981	0.7298
$\widehat{\rho_1\text{-Q}_n}$	0.5345	0.5059	0.5059	0.5286	0.5168	0.5419
$\widehat{\rho_1\text{-T}_n}$	0.6081	0.5893	0.6263	0.6192	0.6252	0.6315

MSE (MEAN SQUARED ERROR) of the estimator over 1000 samples)
(sample size $n = 20$)

estimate \ shape	Tra	Π	LU_{1A}	LU_{1B}	LU_{2A}	LU_{2B}
$\widehat{\rho_2\text{-SD}_n}$	0.0084	0.0086	0.0085	0.0090	0.0085	0.0089
$\widehat{\rho_1\text{-MDD}_n}$	0.0077	0.0080	0.0082	0.0088	0.0080	0.0088
$\widehat{\rho_1\text{-S}_n}$	0.0099	0.0103	0.0104	0.0113	0.0099	0.0111
$\widehat{\rho_1\text{-Q}_n}$	0.0035	0.0037	0.0037	0.0040	0.0036	0.0039
$\widehat{\rho_1\text{-T}_n}$	0.0076	0.0080	0.0083	0.0087	0.0078	0.0086

EXAMPLE OF THE ESTIMATE FOR A SAMPLE
(sample size $n = 100$)

estimate \ shape	Tra	Π	LU_{1A}	LU_{1B}	LU_{2A}	LU_{2B}
$\widehat{\rho_2\text{-SD}_n}$	0.7366	0.7250	0.7515	0.7388	0.7364	0.7448
$\widehat{\rho_1\text{-MDD}_n}$	0.5111	0.4899	0.5039	0.4893	0.4907	0.4792
$\widehat{\rho_1\text{-S}_n}$	0.6572	0.6385	0.6621	0.6796	0.6377	0.6815
$\widehat{\rho_1\text{-Q}_n}$	0.4652	0.4503	0.4673	0.4692	0.4640	0.4712
$\widehat{\rho_1\text{-T}_n}$	0.5652	0.5492	0.5780	0.5840	0.5605	0.5882

MSE (MEAN SQUARED ERROR) of the estimator over 1000 samples)
(sample size $n = 100$)

estimate \ shape	Tra	Π	LU_{1A}	LU_{1B}	LU_{2A}	LU_{2B}
$\widehat{\rho_2\text{-SD}_n}$	0.0015	0.0016	0.0016	0.0016	0.0016	0.0016
$\widehat{\rho_1\text{-MDD}_n}$	0.0015	0.0016	0.0016	0.0018	0.0016	0.0018
$\widehat{\rho_1\text{-S}_n}$	0.0016	0.0016	0.0017	0.0018	0.0016	0.0018
$\widehat{\rho_1\text{-Q}_n}$	0.0004	0.0005	0.0004	0.0005	0.0004	0.0005
$\widehat{\rho_1\text{-T}_n}$	0.0013	0.0014	0.0013	0.0015	0.0013	0.0015

C. Simulating fuzzy outliers

Three different types of outliers are to be considered for all the simulations conducted in this work. We explain now how to generate outliers which will be assumed to be trapezoidal.

First, for each type of outlier, the four-tuple (x_1, x_2, x_3, x_4) is generated from the distribution of the random vector (X_1, X_2, X_3, X_4) in the non-contaminated sample. Then, we construct the outlier $\widetilde{y} = \text{Tra}(y_1 - y_2 - y_3, y_1 - y_2, y_1 + y_2, y_1 + y_2 + y_4)$ in the following way:

- **Outlier of translation:** $y_1 = x_1 + r_1$, $y_2 = x_2$, $y_3 = x_3$, $y_4 = x_4$.
- **Outlier of scale on the core and support:** $y_1 = x_1$, $y_2 = |r_2| \cdot x_2$, $y_3 = |r_2| \cdot x_3$, $y_4 = |r_2| \cdot x_4$.
- **Outlier of both translation and scale:** $y_1 = x_1 + r_1$, $y_2 = |r_2| \cdot x_2$, $y_3 = |r_2| \cdot x_3$, $y_4 = |r_2| \cdot x_4$.

As an example of the process, Figure 4 illustrates on the top the 'symmetric case', generating the non-contaminated sample (in grey) of size 10 from a $\beta(100, 100)$, and on the bottom the 'asymmetric case', generating the non-contaminated sample (in grey) of size 10 from a $\beta(1, 100)$. The chosen weights in both cases have been $\omega_1 = 0.8$, $\omega_2 = 0.1$ and $\omega_3 = 0.1$.

In both cases, two outliers (in black) have been added to the sample. In the symmetric case, for the two outliers of translation we have chosen $r_1 = 30$ and $r_1 = -30$, for the two outliers of scale $r_2 = 5$ and $r_2 = 10$, and for the two outliers of both translation and scale $r_1 = 30$ and $r_2 = 5$ for the first outlier and $r_1 = -30$ and $r_2 = 10$ for the second one. In the asymmetric case, for the two outliers of translation we have chosen $r_1 = 30$ and $r_1 = 60$, for the two outliers of scale $r_2 = 40$ and $r_2 = 80$ and for the two outliers of both translation and scale $r_1 = 30$ and $r_2 = 40$ for the first outlier, and $r_1 = 60$ and $r_2 = 80$ for the second one.

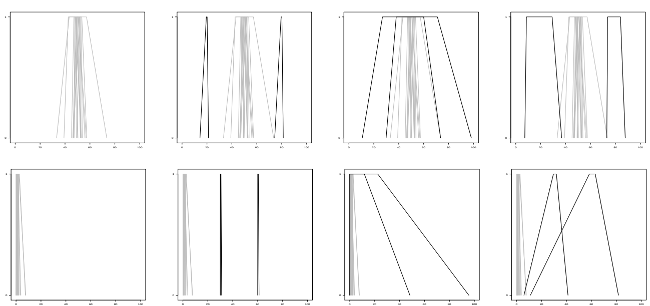


Fig. 4. On the top, from left to right, the non-contaminated sample (in grey) from a symmetric distribution, the contaminated sample by 2 outliers of translation (in black), the contaminated sample by 2 outliers of scale (in black) on core and support and the contaminated sample by 2 outliers (in black) of both location and scale. On the bottom, from left to right, the non-contaminated sample (in grey) from an asymmetric distribution, the contaminated sample by 2 outliers of translation (in black), the contaminated sample by 2 outliers of scale (in black) on core and support and the contaminated sample by 2 outliers (in black) of both location and scale

D. Simulation-based breakdown point approach

It should be noticed that the breakdown point approach does not make proper sense for this realistic simulation procedure, since the supports of the generated fuzzy numbers are assumed to be included within the bounded interval $[0, 100]$. In such a situation, none scale estimator can explode to infinity. Therefore, instead of the finite sample breakdown point, one should better refer to a 'pseudo-breakdown point'.

In the simulations we have considered in this section, the sample size is $n = 21$, and the chosen weights are $\omega_1 = 16/21$, $\omega_2 = 3/21$ and $\omega_3 = 2/21$.

For the ***explosion pseudo-breakdown point***, an outlier of translation $\widetilde{y}_i$ has been constructed so that

– for the symmetric case, the non-contaminated sample has been generated on the basis of a beta distribution $\beta(1000, 1000)$ and the fuzzy numbers have been constrained to belong to the interval $[47.5, 52.5]$. Then, we have chosen

$$r_i^1 = \begin{cases} 8 + \frac{i+1}{2} \cdot 4 & \text{if } i \text{ is odd} \\ -(8 + \frac{i}{2} \cdot 4) & \text{if } i \text{ is even.} \end{cases}$$

– for the asymmetric case, the non-contaminated sample has been generated on the basis of a beta distribution $\beta(1, 100)$ and the fuzzy numbers have been constrained to belong to the interval $[0, 5]$. Then, we have chosen $r_i^1 = 40 + i \cdot 5$.

The general scheme of the simulation of outliers has been structured as follows:

Step 1. A sample $\widetilde{\mathbf{x}}_{21}$ of 21 trapezoidal fuzzy numbers has been simulated from the considered distribution by using the weights $\omega_1 = 16/21$, $\omega_2 = 3/21$ and $\omega_3 = 2/21$.

Step 2. Contaminated samples $\widetilde{\mathbf{y}}_{21,k}$ have been obtained by replacing k observations of the original sample $\widetilde{\mathbf{x}}_{21}$ by k outliers $\widetilde{y}_i$, with $k \in \{1, \ldots, 11\}$ and $i \in \{1, \ldots, k\}$. Overall, k contaminated samples, one for each k value, have been considered.

Step 3. The values of the different location-free scale measures and SD and MDD have been calculated for the original sample without contamination, $\widetilde{\mathbf{x}}_{21}$, and for each of the k contaminated samples $\widetilde{\mathbf{y}}_{21,k}$.

In case any of the 4 real values characterizing an outlier falls outside the reference interval $[0, 100]$, then it is automatically replaced by 0 if it is negative, or by 100 if it is over 100.

The simulation-based conclusions for the explosion in this study are presented in the two first rows in Table V, which display the values of different estimates when outliers are introduced in the sample by replacement.

For the ***implosion pseudo-breakdown point***, the non-contaminated sample has been generated on the basis of a beta distribution $\beta(100, 100)$ regarding the symmetric distribution, and on the basis of a beta distribution $\beta(1, 100)$ regarding the asymmetric distribution.

To study the breakdown point for implosion, we have considered the inliers being all of them equal to one observation chosen randomly from the non-contaminated sample.

The general scheme of the simulation has been structured as follows:

Step 1. A sample $\widetilde{\mathbf{x}}_{21}$ of 21 trapezoidal fuzzy numbers has been simulated from the considered distribution by using the weights $\omega_1 = 16/21$, $\omega_2 = 3/21$ and $\omega_3 = 2/21$.

Step 2. Contaminated samples $\widetilde{\mathbf{y}}_{21,k}$ have been obtained by replacing k observations of the original sample $\widetilde{\mathbf{x}}_{21}$ by k inliers $\widetilde{y}$, with $k \in \{1, \ldots, 20\}$. In total, k contaminated samples, one for each k value.

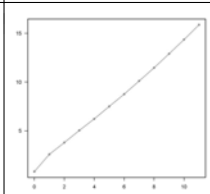
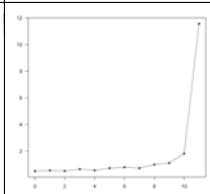
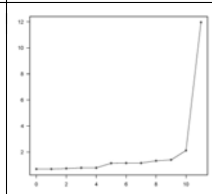
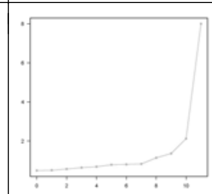
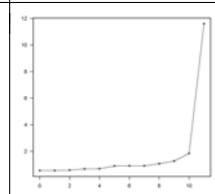
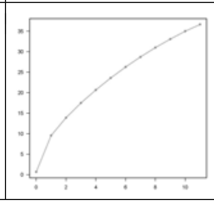
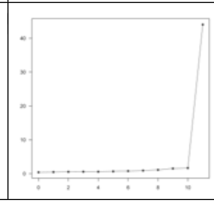
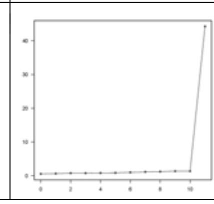
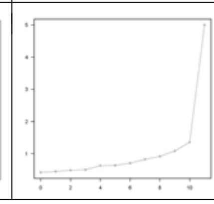
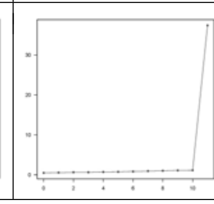
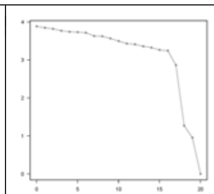
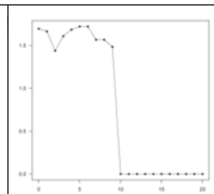
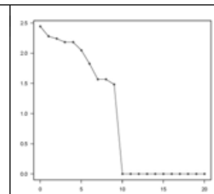
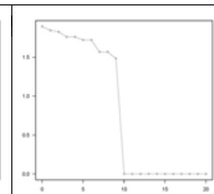
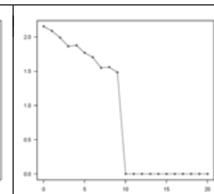
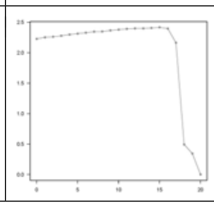
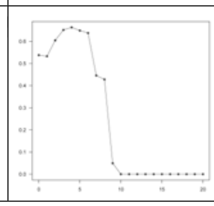
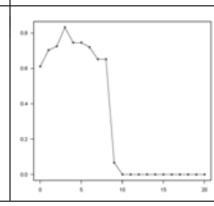
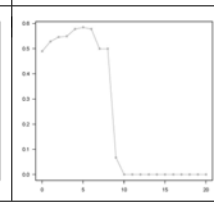
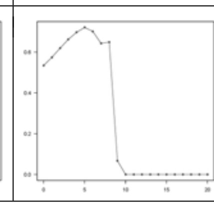
Step 3. The values of the different location-free scale measures and SD and MDD have been calculated for the original sample without contamination, $\widetilde{\mathbf{x}}_{21}$, and for each of k contaminated samples $\widetilde{\mathbf{y}}_{21,k}$.

The simulation-based conclusions for the implosion in this study are presented in the two last rows in Table V, which display the values of different estimates when inliers are introduced in the sample by replacement.

Regarding explosion we can conclude that the minimum number of perturbed observations by outliers that makes the estimator increase noticeably, independently of the considered distribution, has been

- 1 for estimate $\widehat{\rho_2\text{-SD}}(\widetilde{\mathbf{x}}_{21})$,
- 11 for estimates $\widehat{\rho_1\text{-MDD}}(\widetilde{\mathbf{x}}_{21})$, $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_{21})$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_{21})$ and $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_{21})$.

TABLE V
EXPLOSION PSEUDO-BREAKDOWN POINT: VALUES OF THE SCALE ESTIMATES FOR A SAMPLE OF SIZE 21 FROM A SYMMETRIC (1ST ROW) AND ASYMMETRIC (2ND ROW) DISTRIBUTION WITH k (ABSCISE) OBSERVATIONS REPLACED BY OUTLIERS OF TRANSLATION, k VARYING FROM 0 TO 11. IMPLOSION PSEUDO-BREAKDOWN POINT: VALUES OF THE SCALE ESTIMATES FOR A SAMPLE OF SIZE 21 FROM A SYMMETRIC (3RD ROW) AND ASYMMETRIC (4TH ROW) DISTRIBUTION WITH k (ABSCISE) OBSERVATIONS REPLACED BY INLIERS, k VARYING FROM 0 TO 20

PSEUDO-BP	DISTRIBUTION	$\widehat{\rho_2\text{-SD}_n}$	$\widehat{\rho_1\text{-MDD}_n}$	$\widehat{\rho_1\text{-S}_n}$	$\widehat{\rho_1\text{-Q}_n}$	$\widehat{\rho_1\text{-T}_n}$
explosion	symmetric					
explosion	asymmetric					
implosion	symmetric					
implosion	asymmetric					

Regarding implosion we can conclude that the minimum number of perturbed observations by inliers that makes the estimator implode to zero, independently of the distribution case considered, has been

- 20 for estimate $\widehat{\rho_2\text{-SD}}(\widetilde{\mathbf{x}}_{21})$,
- 10 for estimates $\widehat{\rho_1\text{-MDD}}(\widetilde{\mathbf{x}}_{21})$, $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_{21})$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_{21})$ and $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_{21})$.

Therefore, the empirical value for the explosion and implosion 'pseudo-breakdown point' coincides with the value of the theoretical fsbp when it makes proper sense.

Results for other simulation procedures, the other two types of outliers, other sample sizes and other scale estimates can be found in http://bellman.ciencias.uniovi.es/SMIRE/Archivos/SimScest.pdf. Similar conclusions could be drawn for other choices of weights ω_1, ω_2 and ω_3.

E. Simulation-based extended sensitivity curve approach

Another important and useful tool to measure the robustness of an estimator dealing with real-valued data is the sensitivity curve, which represents the sample version of the influence functions (see, for instance, Maronna *et al.* [33] and Rossello [34]).

The finite sample breakdown point (or its 'pseudo' adapted version) tells us how much the estimate changes when a percentage of the data is contaminated by outliers or inliers. In contrast to the finite sample breakdown point, the sensitivity curve describes how the estimator reacts to a single outlier in the data, the outlier being characterized in terms of a real-valued deviation from an arbitrary datum.

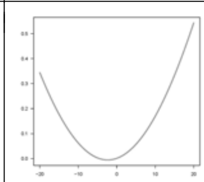
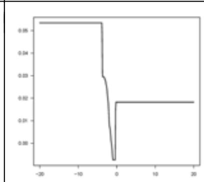
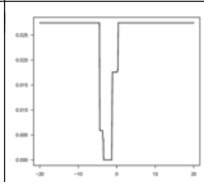
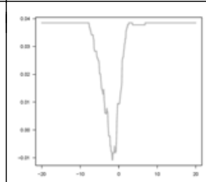
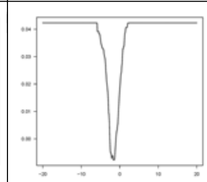
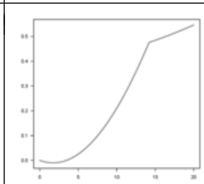
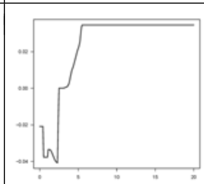
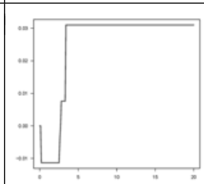
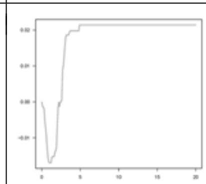
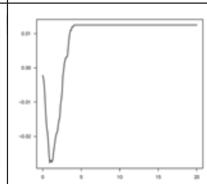
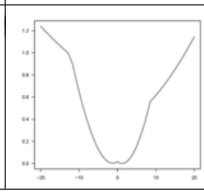
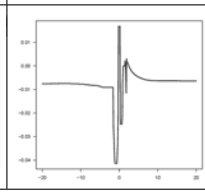
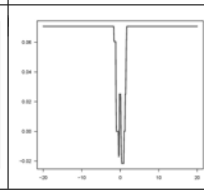
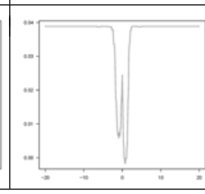
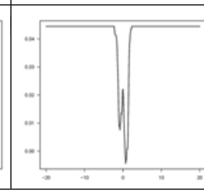
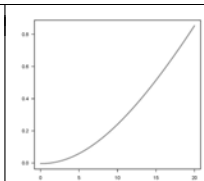
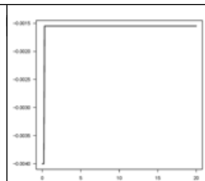
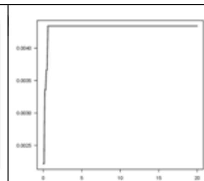
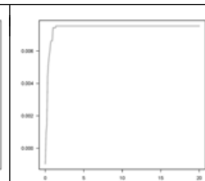
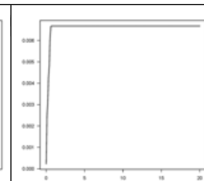
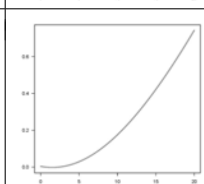
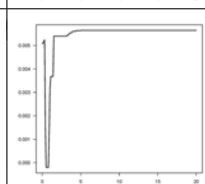
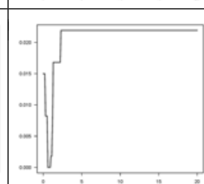
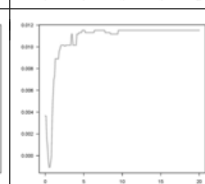
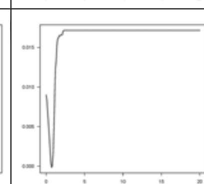
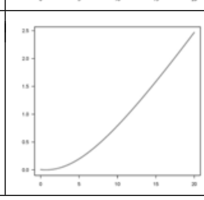
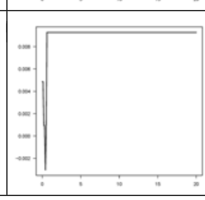
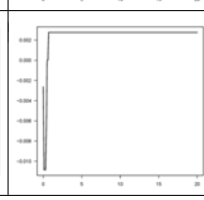
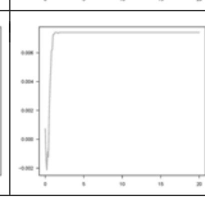
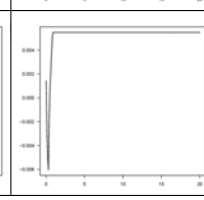
It should be emphasized that in the setting of fuzzy data it would be very complex to extend and graphically display the sensitivity curve. Thus, the above mentioned real-valued deviation would have to be replaced by a fuzzy-valued one, which would be difficult to handle and interpret.

The following analysis concerns a ***real-valued extended sensitivity curve approach***, which corroborates from a different perspective the robustness of the new estimates in this paper. This extended curve is defined so that if $\mathcal{X}$ is an RFN, $\widetilde{\mathbf{x}}_n = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is a sample of observations from $\mathcal{X}$ and $\widehat{\tau}$ is a scale estimate, the **real-valued extended sensitivity curve** of the estimate $\widehat{\tau}(\widetilde{\mathbf{x}}_n)$ is the function associating with each $s \in \mathbb{R}$ (or in a bounded interval, in case the support of the RFN values is constrained to be included in such an interval, where we would refer to pseudo-sensitivity curve) the difference

$$\mathrm{SC}(s) = \widehat{\tau}(\widetilde{\mathbf{x}}_n^{[s]}) - \widehat{\tau}(\widetilde{\mathbf{x}}_n)$$

where the sample $\widetilde{\mathbf{x}}_n^{[s]}$ is obtained by replacing a fixed observation of $\widetilde{\mathbf{x}}_n$, which has been previously chosen at random, by the outlier $\widetilde{y}_s$.

TABLE VI
PSEUDO-SENSITIVITY CURVES OF THE SCALE ESTIMATES AS FUNCTIONS OF s FOR A SAMPLE OF SIZE 100 AND OUTLIERS OF TRANSLATION (1ST/4TH ROWS), SCALE ON CORE AND SUPPORT (2ND/5TH ROWS), AND BOTH (3RD/6TH ROWS) FOR SYMMETRIC/ASYMMETRIC DISTRIBUTIONS

OUTLIERS	DISTRIBUTION	$\widehat{\rho_2\text{-SD}_n}$	$\rho_1\text{-}\widehat{\text{MDD}_n}$	$\widehat{\rho_1\text{-S}_n}$	$\widehat{\rho_1\text{-Q}_n}$	$\widehat{\rho_1\text{-T}_n}$
translation	symmetric					
scale core-support	symmetric					
transl.+scale core-supp.	symmetric					
translation	asymmetric					
scale core-support	asymmetric					
transl.+scale core-supp.	asymmetric					

In the analysis in this section, the non-contaminated sample has been simulated in accordance with the above described procedure. We have considered the sample size to be $n = 100$, with the weights $\omega_1 = 0.8, \omega_2 = 0.1$ and $\omega_3 = 0.1$. The outlier $\widetilde{y}_s$ has been constructed as follows:

- For the symmetric case, the non-contaminated sample has been generated from a beta $\beta(100, 100)$, and
  - **Outlier of translation:** $r_s^1 = s$, with s varying from -20 to 20 with a step equal to 0.1.
  - **Outlier of scale on the core and support:** $r_s^2 = s$, with s varying from 0 to 20 with a step equal to 0.1.
  - **Outlier of both translation and scale:** $r_s^1 = r_s^2 = s$, with s varying from -20 to 20 with a step equal to 0.1.
- For the asymmetric case, the non-contaminated sample has been generated from a beta $\beta(1, 100)$.
  - **Outlier of translation:** $r_s^1 = s$, with s varying from 0 to 20 with a step equal to 0.1.
  - **Outlier of scale on the core and support:** $r_s^2 = s$, with s varying from 0 to 20 with a step equal to 0.1.
  - **Outlier of both translation and scale:** $r_s^1 = r_s^2 = s$, with s varying from 0 to 20 with a step equal to 0.1.

For each type of outlier, the general scheme of the construction of the pseudo-sensitivity curves has been as follows:

Step 1. A sample $\widetilde{\mathbf{x}}_{100}$ of 100 trapezoidal fuzzy numbers has been simulated from the considered distribution.

Step 2. One observation from the original sample $\widetilde{\mathbf{x}}_{100}$ has been chosen randomly and replaced by the outlier $\widetilde{y}_s$.

Step 3. For each s, the value of the sensitivity curve has been calculated for each estimator of scale.

The pseudo-sensitivity curves have been graphically displayed for each estimator in Table VI.

Irrespective of the type of outlier we are considering, the pseudo-sensitivity curves

- show an increasing behaviour w.r.t. $|s|$ for $\widehat{\rho_2\text{-SD}}(\widetilde{\mathbf{x}}_n)$;
- show an upper bounded and very similar behaviour w.r.t. $|s|$ for $\widehat{\rho_1\text{-MDD}}(\widetilde{\mathbf{x}}_n)$, $\widehat{\rho_1\text{-S}}(\widetilde{\mathbf{x}}_n)$, $\widehat{\rho_1\text{-Q}}(\widetilde{\mathbf{x}}_n)$ and $\widehat{\rho_1\text{-T}}(\widetilde{\mathbf{x}}_n)$.

As for the fsbp, one can also find in the link http://bellman.ciencias.uniovi.es/SMIRE/Archivos/SimScest.pdf the results for other simulation procedures, the other two types of outliers, other sample sizes and other scale estimates. Similar conclusions could be drawn for other choices of weights ω_1, ω_2 and ω_3. In all cases, we can see that the sensitivity curves for the $\widehat{\rho_1\text{-T}_n}$ are smoother than for the rest of robust scale estimates, and the sensitivity curves for non-robust estimates show an increasing behaviour w.r.t. $|s|$.

VI. Concluding remarks

The definition of location-free robust scale estimates for fuzzy data aims to substantially ease the computation of location-based robust scale estimates for fuzzy data, like the Median Distance Deviation, which has recently been introduced (see [8]). Robust location measures for fuzzy data do not often preserve the shape of sample data, and to involve them in measuring scale frequently becomes a cumbersome and not fully exact task. To avoid such drawbacks, three location-free scale estimates are introduced and examined in this paper. They extend the so-called explicit scale estimates for numerical data proposed by Rousseeuw and Croux (see [23], [24]). Fortunately, the extension preserves the key properties of the estimates for numerical data and inherits their robustness (extended proofs for this are included in the supplementary material). In addition to the theoretical proofs and the illustrative simulations in connection with the finite sample breakdown point for the alternate new scale estimates, the paper shows for the first time simulation-based sensitivity studies for the location-based and location-free scale estimates w.r.t. outliers of translation and outliers of scale on core and support for fuzzy datasets.

The introduced estimates fulfill the following properties:

- Their finite sample breakdown point equals $\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor$ (like MDD), which is the highest fsbp a scale estimate can reach.
- They have a simple and explicit formula.
- They are location-free estimates of scale, in contrast to the estimates $\widehat{\rho_1\text{-MDD}_n}$ (or $\widehat{\rho_2\text{-SD}_n}$) which are based on location measures (the 1-norm median and Aumann-type mean, respectively).
- Their sensitivity curves are upper bounded (like MDD).

Furthermore, whereas the computation of the MDD for fuzzy data should be generally performed approximately and requires important computational time cost and complexity, the computation of the new estimates is much simpler for the usual types of fuzzy data and very easy-to-make for trapezoidal ones. Actually, their computation for trapezoidal fuzzy data is scarcely more time consuming and complex than the computation with real-valued data (see the R package [35] implementing such computations).

In summary, the scale estimates for fuzzy data in this paper share the main statistical properties of the MDD and show a similar behavior, but *in the setting of fuzzy data location-freedom entails an important added value in estimating scale* since it considerably eases computations.

It should be remarked that since there are not realistic models for the normality of RFNs, one cannot formally analyze the Gaussian efficiency of the new estimates in contrast to that of the MDD. If one makes use of the normality by Puri and Ralescu [36], the conclusions will be certainly similar to those for the real-valued case, but such a normal model is quite restrictive and not realistic enough.

Finally, among the future directions to be considered aiming to complete and extend the study carried out in this paper, one can mention:

- an extension and comparative analysis of M-estimates of scale;
- the definition and comparative analysis of the median distance deviation about the M-estimator of location (see Sinova *et al.* [5]);
- an analysis of the influence of the property of symmetry of an RFN (see Sinova *et al.* [37]) in some properties of the scale estimates.

Acknowledgements

The authors are deeply grateful to the Associate Editor and reviewers for their very helpful suggestions.

References

[1] M.L. Puri and D.A. Ralescu, “Fuzzy random variables,” *J. Math. Anal. Appl.*, Vol. 114, pp. 409–422, 1986.

[2] B. Sinova, M.A. Gil, A. Colubi and S. Van Aelst, “The median of a random fuzzy number. The 1-norm distance approach,” *Fuzzy Sets Syst.*, Vol. 200, pp. 99-115, 2012.

[3] B. Sinova, S. de la Rosa de Sáa and M.A. Gil, “A generalized L^1-type metric between fuzzy numbers for an approach to central tendency of fuzzy data,” *Inform. Sci.*, Vol. 242, pp. 22–34, 2013.

[4] B. Sinova, S. Pérez-Fernández and M. Montenegro, “The wabl/ldev/rdev median of a random fuzzy number and statistical properties,” in: *Strengthening Links Between Data Analysis and Soft Computing* (P. Grzegorzewski, M. Gagolewski, O. Hryniewicz and M.A. Gil, Eds.). Springer: Heidelberg, 2015, pp. 143–150.

[5] B. Sinova, M.A. Gil and S. Van Aelst, “M-estimates of location for the robust central tendency of fuzzy data,” *IEEE Trans. Fuzzy Syst.* Vol. 24, pp. 945–956, 2016.

[6] R. Körner, “On the variance of fuzzy random variables,” *Fuzzy Set. Syst.*, Vol. 92, pp. 83–93, 1997.

[7] M.A. Lubiano, M.A. Gil, M. López-Díaz and M.T. López, “The $\vec{\boldsymbol{\lambda}}$-mean squared dispersion associated with a fuzzy random variable,” *Fuzzy Set. Syst.*, Vol. 111, pp. 307–317, 2000.

[8] S. de la Rosa de Sáa, M.A. Lubiano, B. Sinova and P. Filzmoser, “Robust scale estimators for fuzzy data,” *Adv. Data Anal. Classif.*, Vol. 11, pp. 731–758, 2017.

[9] L.A. Zadeh, “The concept of a linguistic variable and its application to approximate reasoning. Part 1,” *Inform. Sci.*, Vol. 8, pp. 199–249; “Part 2,” *Inform. Sci.*, Vol. 8, pp. 301–353; “Part 3,” *Inform. Sci.*, Vol. 9, pp. 43–80, 1975.

[10] P. Diamond and P. Kloeden, “Metric spaces of fuzzy sets,” *Fuzzy Sets Syst.*, Vol. 35, pp. 241–249, 1990.

[11] M.A. Lubiano, A. Salas and M.A. Gil, “A hypothesis testing-based discussion on the sensitivity of means of fuzzy data with respect to data shape,” *Fuzzy Sets Syst.*, Vol. 328, pp. 54–69, 2017.

[12] V. Krätschmer, “Integrals of random fuzzy sets,” *Test*, Vol. 15, pp. 433–469, 2006.

[13] S. de la Rosa de Sáa, M.A. Gil, G. González-Rodríguez, M.T. López and M.A. Lubiano, "Fuzzy rating scale-based questionnaires and their statistical analysis," *IEEE Trans. Fuzzy Syst.*, Vol. 23, pp. 111–126, 2015.

[14] M. Kaluszka and M. Krzeszowiec, "On risk aversion under fuzzy random data," *Fuzzy Sets Syst.*, Vol. 328, pp. 35–53, 2017.

[15] Y. Ogura, S. Li and X. Wang, "Large and moderate deviations of random upper semicontinuous functions," *Stoch. Anal. Appl.*, Vol. 28, pp. 350–376, 2010.

[16] N.V. Quang and P.T. Nguyen, "Some strong laws of large number for double array of random upper semicontinuous functions in convex combination spaces," *Statist. Prob. Lett.*, Vol. 96, pp. 85–94, 2015.

[17] N.T. Thuan and N.V. Quang, "Negative association and negative dependence for random upper semicontinuous functions, with applications," *J. Multiv. Anal.*, Vol. 145, pp. 44–57, 2016.

[18] D. Dubois and H. Prade, "Additions of interactive fuzzy numbers," *IEEE Trans. Automat. Control*, Vol. 26, pp. 926–936, 1981.

[19] P.J. Bickel and E.L. Lehmann, "Descriptive statistics for nonparametric models III: dispersion," *Ann. Statist.*, Vol. 4, pp. 1139–1158, 1976.

[20] P.J. Bickel and E.L. Lehmann, "Descriptive statistics for nonparametric models IV: spread," in: *Contributions to Statistics, Hájek Memorial Volume* (J.Jurečková, Ed.). Prague: Academia, 1976, pp. 33–40.

[21] H. Lee-Kwang and J.-H. Lee, "A method for ranking fuzzy numbers and its application to decision-making," *IEEE Trans. Fuzzy Syst.*, Vol. 7, pp. 677–685, 1999.

[22] A.I. Ban and L. Coroianu, "Simplifying the search for effective ranking of fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, Vol. 23, pp. 327–339, 2015.

[23] P.J. Rousseeuw and C. Croux, "Explicit scale estimators with high breakdown point," in: *L_1-Statistical Analysis and Related Methods* (Y. Dodge, Ed). Amsterdam: North-Holland, 1992, pp. 77–92.

[24] P.J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," *J. Amer. Statist. Assoc.*, Vol. 88, pp. 1273–1283, 1993.

[25] T. Hesketh, R. Pryor and B. Hesketh, " An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences," *Int. J. Man-Mach. Stud.*, Vol. 29, pp. 21–35, 1988.

[26] S. de la Rosa Sáa, C. Carleos, M.T. López and M. Montenegro, "A case study-based analysis of the influence of the fuzzy data shape in quantifying their Fréchet's variance," in: *The Mathematics of the Uncertain: A Tribute to Pedro Gil* (E. Gil, E. Gil, J. Gil and M.A. Gil, Eds.). Cham: Springer, 2018, pp. 663–672.

[27] L. Stefanini, L. Sorini and M.L. Guerra, "Parametric representation of fuzzy numbers and applications to fuzzy calculus," *Fuzzy Sets Syst.*, Vol. 157, pp. 2423–2455, 2006.

[28] D.L. Donoho and P.J. Huber, "The notion of breakdown point," in: *A Festschrift for Erich L. Lehmann* (P.J. Bickel, K. Doksum and J.L. Hodges Jr., Eds.). Belmont: Wadsworth, 1983, pp. 157–184.

[29] F.R. Hampel, *Contributions to the Theory of Robust Estimation*. PhD Thesis, University of California at Berkeley, 1968.

[30] D.L. Donoho, *Breakdown properties of multivariate location estimators*. PhD qualifying paper, Dept. Statistics, Harvard University, 1982.

[31] M.A. Lubiano, A. Salas, C. Carleos, S. de la Rosa de Sáa and M.A. Gil, "Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data," *Int. J. Approx. Reas.*, Vol. 88, pp. 128–147, 2017.

[32] M.A. Lubiano, S. De la Rosa de Sáa, B. Sinova and M.A. Gil "Empirical sensitivity analysis on the influence of the shape of fuzzy data on the estimation of some statistical measures," in: *Strengthening Links Between Data Analysis and Soft Computing* (P. Grzegorzewski, M. Gagolewski, O. Hryniewicz and M.A. Gil, Eds.). Springer: Heidelberg, 2015, pp. 123–131.

[33] R.A. Maronna, R.D. Martin and V.J. Yohai, *Robust Statistics: Theory and Methods*. Chinchester: J. Wiley & Sons, 2006.

[34] D. Rossello, "Ranking of investment funds: acceptability versus robustness," *Eur. J. Oper. Res.*, Vol. 245, pp. 828–836, 2015.

[35] M.A. Lubiano and S. de la Rosa de Sáa, *FuzzyStatTra: Statistical Methods for Trapezoidal Fuzzy Numbers*. https://cran.r-project.org/web/packages/FuzzyStatTra/index.html.

[36] M.L. Puri and D.A. Ralescu, "The concept of normality for fuzzy random variables," *Ann. Probab.*, Vol. 11, pp. 1373–1379, 1985.

[37] B. Sinova, M.R. Casals and M.A. Gil, "Central tendency for symmetric random fuzzy numbers," *Inform. Sci.*, Vol. 278, pp. 599–613, 2014.

Sara de la Rosa de Sáa received her BSc and MSc degrees in Mathematics in 2011 and 2012, respectively, from the University of Oviedo, Spain. In 2017, she has received a Joint PhD degree in Mathematics and Statistics from the Universities of Oviedo and TU Wien. Her current research interests include the development of new statistical techniques for the analysis of fuzzy data, as well as their application to real-life problems and the discussion on the potential benefits of using these procedures. She is involved in some research projects regarding the statistical analysis of biomedical data.

María Asunción Lubiano received her BSc and MSc degrees in Mathematics from the University of Valladolid and her PhD degree (also in Maths) from the University of Oviedo. Since 2001 she is associate professor with the Department of Statistics, O.R. and D.M. of the University of Oviedo. She has participated in several research projects and some R+D contracts with companies. She has coedited three international books in connection with international conferences on Soft Methods in Probability and Statistics, and she was the Technical Editor of the statistical journal *TEST* (2005-2008). She is Area Editor of the *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. Her research interests are (amongst others) the statistical analysis and especially the study of the variability associated with imprecise random elements. For the last years she is one of the two main responsible of the development of a software for the analysis of fuzzy data with the statistical package R.

Beatriz Sinova received her BSc and MSc degrees in Mathematics in 2009 and 2010, respectively, from the University of Oviedo, Spain. In 2014, she has received a Joint PhD degree in Mathematics and Statistics from the Universities of Oviedo and Gent. Her current research interests include the development of new statistical techniques for the location of imprecise-valued data, as well as their application to real-life problems. She is involved in some research projects regarding the statistical analysis of interval-, fuzzy- and functional-valued data. She has been awarded with the Ramiro Melendreras 2016 Prize of the Spanish Society of Statistics and Operational Research and with the Vicent Caselles 2017 Prize of the Royal Spanish Society of Mathematics and the BBVA Foundation.

María Ángeles Gil is full professor of Statistics and O.R. at the University of Oviedo (Spain) since 1992. Her research was initially focused on the measurement and statistical developments concerning the amount of information and inequality of random variables. Since 1983, her research interests have been mostly related to the statistical analysis of imprecise data (mainly set-valued and fuzzy set-valued). On this topic she heads the research group SMIRE (http://bellman.ciencias.uniovi.es) and she is a member of the Core SMPS Group steering the biennial conferences on Soft Methods in Probability and Statistics. She was co-editor-in-chief of the journal *TEST* (2005-2008) and associate/area editor for *Computational Statistics and Data Analysis*, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, and *TEST* and she is currently Associate Editor for *Fuzzy Sets and Systems*, *International Transactions in Operational Research*, and *Information Sciences*. She has been awarded with the Silver Medal of the Principality of Asturias in 2014 and is an IFSA fellow since 2015.

Peter Filzmoser is full professor at the statistics department of the Vienna University of Technology. His main research interests include robust statistics, methods for compositional data analysis, statistical computing with R, and many more. He is author of the books: "Applied Compositional Data Analysis. With Worked Examples in R (Springer 2018), "Introduction to Multivariate Statistical Analysis in Chemometrics" (CRC Press, 2009), and "Statistical Data Analysis Explained. Applied Environmental statistics with R (Wiley, 2008). Currently, he is head of the Research Unit in Computational Statistics, Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology.