Tukey's biweight loss function for fuzzy set-valued M-estimators of location

Beatriz Sinova[1,3] and Stefan Van Aelst[2,3]

Abstract The Aumann-type mean is probably the best-known measure for the location of a random fuzzy set. Despite its numerous probabilistic and statistical properties, it inherits from the mean of a real-valued random variable the high sensitivity to outliers or data changes. Several alternatives extending the concept of median to the fuzzy setting have already been proposed in the literature. Recently, the adaptation of location M-estimators has also been tackled. The expression of fuzzy-valued location M-estimators as weighted means under mild conditions allows us to guarantee that these measures take values in the space of fuzzy sets. It has already been shown that these conditions hold for the Huber and Hampel families of loss functions. In this paper, the strong consistency and the maximum finite sample breakdown point when the Tukey biweight (or bisquare) loss function is chosen are analyzed. Finally, a real-life example will illustrate the influence of the choice of the loss function on the outputs.

Key words: random fuzzy set, robustness, location M-estimator, bisquare loss function, biweight loss function

Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain
`sinovabeatriz@uniovi.es` ·
Department of Mathematics, KU Leuven, 3001 Leuven, Belgium
`stefan.vanaelst@kuleuven.be` ·
Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Gent, Belgium

1 Introduction

Random fuzzy sets (fuzzy random variables in Puri and Ralescu's sense [10]) are an appropriate mathematical model to formalize numerous real-life experiments characterized by an underlying imprecision. In order to analyze them statistically, a wide range of methods has been proposed during the last years. Unfortunately, most of this methodology is based on the Aumann-type mean, which is a well-known location measure for random fuzzy sets that fulfills many convenient properties from both the statistical and probabilistic points of view, but it presents a high sensitivity to outliers or data changes. With the aim of providing a more robust central tendency measure, several extensions of the concept of median have already been published. However, this paper focuses on the more recent and more general M-estimation approach.

Kim and Scott [9] have studied M-estimators in the kernel density estimation context, but their theory remains valid for Hilbert-valued random elements. The space of fuzzy sets can be isometrically embedded into a convex cone of a Hilbert space, which allowed us to adapt some of their results to the fuzzy-valued case in Sinova *et al.* [12]. Although only the one-dimensional case (random fuzzy numbers) has been specified in [12], location M-estimators can be analogously defined for random fuzzy sets and studied as in this paper.

Sufficient conditions are provided in Sinova *et al.* [12] to guarantee that the adaptation of Kim and Scott's results is valid, that is, that location M-estimators belong to the convex cone of the Hilbert space. Among the loss functions satisfying such assumptions, Huber's and Hampel's loss functions were analyzed in [12] to prove the strong consistency of the corresponding M-estimators and show that the maximum finite sample breakdown point is attained. Another well-known family of loss functions, Tukey's biweight (also referred to as the bisquare function), is considered in this paper. Apart from checking that the sufficient conditions also hold for this choice, the strong consistency of the Tukey location M-estimator is established and its finite sample breakdown point is derived. Proofs are based on the same sketches included for the one-dimensional case in Sinova *et al.* [12].

In Section 2, location M-estimators for random fuzzy sets are introduced and the Representer Theorem, which expresses them as weighted means under certain sufficient conditions, is recalled. In Section 3, the choice of Tukey's biweight loss function is analyzed in terms of the strong consistency of the resulting estimator and its finite sample breakdown point. A real-life example in Section 4 illustrates the influence of the choice of the loss function on the outputs. Finally, some concluding remarks are provided in Section 5.

2 Location M-estimators for random fuzzy sets

In this section, location M-estimators are adapted to summarize the central tendency of random fuzzy sets. M-estimation, firstly introduced by Huber [7], is a well-established approach that yields robust estimators. The key idea behind them is to restrict the influence of outliers by substituting the square of "errors" in methods like least squared and maximum likelihood for a (usually less rapidly increasing) loss function applied to the errors of the data. The loss function, denoted by ρ, is usually assumed to vanish at 0 and to be even and non-decreasing for positive values.

Let $p \in \mathbb{N}$, $\mathcal{F}_c^*(\mathbb{R}^p)$ denote the space of bounded fuzzy sets and D represent a metric defined on $\mathcal{F}_c^*(\mathbb{R}^p) \times \mathcal{F}_c^*(\mathbb{R}^p)$ whose associated norm fulfills the parallelogram law (which allows the isometrical embedding of $\mathcal{F}_c^*(\mathbb{R}^p)$ into the convex cone of a Hilbert space).

Definition 1. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R}^p)$ an associated random fuzzy set. Moreover, let ρ be a continuous loss function, and $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ a simple random sample from $\mathcal{X}$. Then, the **fuzzy M-estimator of location** is the fuzzy set-valued statistic $\widehat{\widetilde{g}^M}[(\mathcal{X}_1, \ldots, \mathcal{X}_n)]$, given, if it exists, by

$$\widehat{\widetilde{g}^M}[(\mathcal{X}_1, \ldots, \mathcal{X}_n)] = \arg \min_{\widetilde{g} \in \mathcal{F}_c^*(\mathbb{R}^p)} \frac{1}{n} \sum_{i=1}^{n} \rho(D(\mathcal{X}_i, \widetilde{g})).$$

Now, a result by Kim and Scott [9] is adapted to the fuzzy-valued case. The Representer Theorem (Theorem 1) is crucial for the particularization of Kim and Scott's theory about M-estimation for the kernel density estimation problem to random fuzzy sets. The conditions they assume to ensure the existence of M-estimates of location allow us to express the M-estimates as weighted means of the sample elements and, consequently, to assure that the M-estimates are indeed fuzzy set-valued statistics.

Theorem 1. *Consider the metric space $(\mathcal{F}_c^*(\mathbb{R}^p), D)$. Let $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ be a simple random sample from a random fuzzy set $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R}^p)$ on a probability space $(\Omega, \mathcal{A}, P)$. Moreover, let ρ be a continuous loss function which satisfies the assumptions*

- *ρ is non-decreasing for positive values, $\rho(0) = 0$ and $\lim_{x \to 0} \rho(x)/x = 0$,*
- *Let $\phi(x) = \rho'(x)/x$ and $\phi(0) \equiv \lim_{x \to 0} \phi(x)$, assuming that $\phi(0)$ exists and is finite.*

Then, the M-estimator of location exists and it can be expressed as

$$\widehat{\widetilde{g}^M}[(\mathcal{X}_1, \ldots, \mathcal{X}_n)] = \sum_{i=1}^{n} \omega_i \cdot \mathcal{X}_i$$

with $\omega_i \geq 0$, $\sum_{i=1}^{n} \omega_i = 1$. Furthermore, $\omega_i \propto \phi(D(\mathcal{X}_i, \widehat{\widetilde{g}^M}[(\mathcal{X}_1, \ldots, \mathcal{X}_n)]))$.

In Sinova *et al.* [12], the well-known Huber and Hampel families of loss functions were used to compute M-estimators. Recall that the *Huber loss function* [8] is given by

$$\rho_a^H(x) = \begin{cases} x^2/2 & \text{if } |x| \leq a \\ a(|x| - a/2) & \text{otherwise,} \end{cases}$$

with $a > 0$ a tuning parameter, while the *Hampel loss function* [5] corresponds to

$$\rho_{a,b,c}(x) = \begin{cases} x^2/2 & \text{if } |x| < a \\ a(|x| - a/2) & \text{if } a \leq |x| < b \\ \dfrac{a(|x| - c)^2}{2(b-c)} + \dfrac{a(b+c-a)}{2} & \text{if } b \leq |x| < c \\ \dfrac{a(b+c-a)}{2} & \text{if } c \leq |x|, \end{cases}$$

where the nonnegative parameters $a < b < c$ allow us to control the degree of suppression of large errors. The smaller their values, the higher this degree. Note that the Huber loss function is convex and puts less emphasis on large errors compared to the squared error loss. On the other hand, Hampel's loss function is not convex and can better cope with extreme outliers, since observations far from the center ($|x| \geq c$) always contribute in the same way to the loss.

Another well-known family of loss functions is the *Tukey biweight or bisquare* [1], given by:

$$\rho_c^T(x) = \begin{cases} c^2/6 \cdot (1 - (1 - (x/c)^2)^3) & \text{if } |x| \leq c \\ c^2/6 & \text{otherwise,} \end{cases}$$

with tuning parameter $c > 0$. This loss function shares with Hampel's one that it is not convex anymore and the contribution of large errors ($|x| \geq c$) to the loss does not change anymore. Therefore, the benefit of the Tukey loss function is to combine the better performance of Hampel's loss function regarding extreme outliers with the simplicity of an expression depending on just one tuning parameter, like the Huber loss function.

It can be easily checked that the family ρ_c^T of loss functions fulfills all the required conditions: they are differentiable, non-decreasing for positive values and even, they vanish at 0, $\lim_{x\to 0} \rho_c^T(x)/x = 0$, $\phi_c^T(0) \equiv \lim_{x\to 0} \phi_c^T(x)$ exists and is finite.

Therefore, all the properties derived from the Representer Theorem in Sinova *et al.* [12] also hold when the Tukey biweight loss function is chosen. In particular, it can be highlighted that Tukey M-estimators of location are translation equivariant, but not scale equivariant in general. With the aim of avoiding the excessive influence of the measurement units on the outputs, due to the lack of scale equivariance unless ρ is a power function, the tuning parameters will be selected based on the distribution of the distances to the center. That is, we first compute an initial robust estimator of location (e.g., the impartial trimmed mean as in Colubi and González-Rodríguez [2] or, if $p = 1$, the 1-norm median in Sinova *et al.* [11]) and then, the distances between each observation and this initial estimate are calculated. Our recommendation is to use the 1-norm median as initial estimate when analyzing random fuzzy numbers, since its computation is not complex and this measure does not depend on the existence or not of outliers in the sample to provide us with a good initial estimate. The impartial trimmed mean (see Colubi and González-Rodríguez [2]) presents the disadvantage of requiring to fix the trimming proportion "a priori" and, in case there are no outliers, the initial estimate could be a bit far from the real center of the sample distribution. The choice for the tuning parameters a, b and c will be, along this paper, the median, the 75$^{\text{th}}$ and the 85$^{\text{th}}$ percentiles of those distances, following Kim and Scott's suggestion [9].

Regarding the practical computation of Tukey M-estimators of location, recall that the standard iteratively re-weighted least squares algorithm (see, for example, Huber [7]) can provide us with an approximation as in [12]:

Step 1. Select initial weights $\omega_i^{(0)} \in \mathbb{R}$, for $i \in \{1, \ldots, n\}$, such that $\omega_i^{(0)} \geq 0$ and $\sum_{i=1}^n \omega_i^{(0)} = 1$ (which is equivalent to choose a robust estimator of location to initialize the algorithm).

Step 2. Generate a sequence $\{\widetilde{g}^M_{(k)}\}_{k\in\mathbb{N}}$ by iterating the following procedure:

$$\widetilde{g}^M_{(k)} = \sum_{i=1}^n \omega_i^{(k-1)} \mathcal{X}_i, \quad \omega_i^{(k)} = \frac{\phi_c^T(D(\mathcal{X}_i, \widetilde{g}^M_{(k)}))}{\sum_{j=1}^n \phi_c^T(D(\mathcal{X}_j, \widetilde{g}^M_{(k)}))}.$$

Step 3. Terminate the algorithm when

$$\frac{|\frac{1}{n}\sum_{i=1}^n \rho_c^T(D(\mathcal{X}_i, \widetilde{g}^M_{(k+1)})) - \frac{1}{n}\sum_{i=1}^n \rho_c^T(D(\mathcal{X}_i, \widetilde{g}^M_{(k)}))|}{\frac{1}{n}\sum_{i=1}^n \rho_c^T(D(\mathcal{X}_i, \widetilde{g}^M_{(k)}))} < \varepsilon,$$

for some desired tolerance $\varepsilon > 0$.

3 Specific properties of fuzzy-valued location M-estimators based on Tukey biweight loss function

The strong consistency of fuzzy number-valued M-estimators of location was studied in Sinova *et al.* [12] for specific loss functions: ρ being either nondecreasing for positive values, subadditive and unbounded or the Huber or Hampel loss function (independently of the values of the tuning parameters).

However, this result can be generalized to cover any bounded loss function and, in consequence, the Tukey biweight choice.

Theorem 2. *Consider the metric space $(\mathcal{F}_c(A), D)$, with A a non-empty compact convex set of $\mathbb{R}^p$ and D topologically equivalent to the mid/spr-based L^2 distance D_θ^ℓ (see Trutschnig* et al. *[13] for details concerning this metric). Let $\mathcal{X} : \Omega \to \mathcal{F}_c(A)$ be a random fuzzy set associated with a probability space $(\Omega, \mathcal{A}, P)$. Under any of the following assumptions:*

- *ρ is non-decreasing for positive values, subadditive and unbounded,*
- *ρ, for positive values, has linear upper and lower bounds with the same slope,*
- *ρ is bounded,*

and whenever the associated M-location value

$$\tilde{g}^M(\mathcal{X}) = \arg\min_{\widetilde{U} \in \mathcal{F}_c(A)} E\left[\rho\left(D(\mathcal{X}, \widetilde{U})\right)\right]$$

exists and is unique, the M-estimator of location is a strongly consistent estimator of $\tilde{g}^M(\mathcal{X})$, i.e.,

$$\lim_{n\to\infty} D(\widehat{\widetilde{g}^M}[(\mathcal{X}_1, \ldots, \mathcal{X}_n)], \tilde{g}^M(\mathcal{X})) = 0 \quad \textit{a.s. [P].}$$

It should be clarified that it is very common in practice to fix a bounded referential, as is the case for the fuzzy rating scale (see Hesketh *et al.* [6]) when $p = 1$.

With respect to the robustness of the location M-estimators based on the Tukey biweight loss function, their *finite sample breakdown point*, for short fsbp (Hampel [4], Donoho and Huber [3]) has been computed. The fsbp represents the smallest fraction of sample observations that needs to be perturbed to make the distances between the original and the contaminated M-estimates arbitrarily large.

Theorem 3. *Consider the metric space $(\mathcal{F}_c^*(\mathbb{R}^p), D)$. Let $\mathcal{X} : \Omega \to \mathcal{F}_c^*(\mathbb{R}^p)$ be a random fuzzy set associated with a probability space $(\Omega, \mathcal{A}, P)$ and let $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ be a sample obtained from $\mathcal{X}$. Moreover, let ρ be a continuous loss function fulfilling the assumptions in Theorem 1, upper bounded by certain $C < \infty$ and satisfying*

$$\rho\left(\max_{1\le i,j\le n} D(\widetilde{x}_i, \widetilde{x}_j)\right) < \frac{n - 2\lfloor\frac{n-1}{2}\rfloor}{n - \lfloor\frac{n-1}{2}\rfloor - 1} \cdot C,$$

and such that the corresponding sample M-estimate of location is unique. Then the finite sample breakdown point of the corresponding location M-estimator is exactly $\frac{1}{n}\lfloor\frac{n+1}{2}\rfloor$, where $\lfloor\cdot\rfloor$ denotes the floor function.

4 Real-life example

A real-life example now illustrates fuzzy-valued location M-estimators.

Example. 68 fourth grade students from Colegio San Ignacio (Oviedo, Spain) have been asked to answer some questions from the joint Student questionnaire TIMSS (Trends in International Mathematics and Science Study) - PIRLS (Progress in International Reading Literacy Study) survey using a fuzzy rating scale (Hesketh *et al.* [6]). To simplify the instructions given to the nine-and-ten-year-old students, only trapezoidal fuzzy numbers have been considered. This study is going to be limited to the item that represents the degree of agreement with the statement "studying mathematics is harder than any other subject".

Location M-estimators based on Huber, Hampel and Tukey loss functions have been computed using the mid/spr-based L^2 distance $D^{\ell}_{\theta=1/3}$, where ℓ denotes the Lebesgue measure on $[0,1]$ (see Trutschnig *et al.* [13]). The 1-norm median in [11] has been considered as the initial robust estimator for the selection of the tuning parameters and the initialization of the algorithm to approximate the M-estimates.

The outputs for the three M-estimates have been displayed in Figure 1.

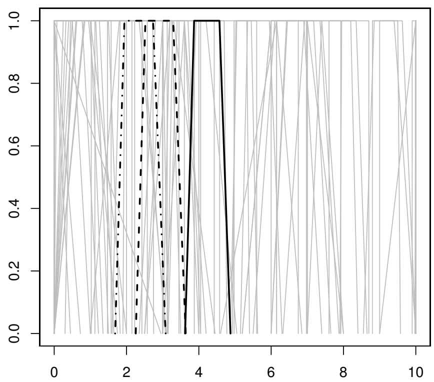

Fig. 1 In black, Huber (solid line), Hampel (dashed line) and Tukey (dash-dot line) M-estimates for the fuzzy-valued data (in grey) from Example

As shown in Sinova *et al.* [12], when analyzing trapezoidal fuzzy numbers, any loss function fulfilling the conditions stated for the Representer Theorem provides us with an M-estimate of trapezoidal shape too.

Notice that the aim of this example is just to illustrate the computation of fuzzy-valued M-estimators and the influence the choice of the loss function has on the outputs, but not to provide a comparison of the different loss functions. On one hand, there are no outliers in the answers given by the students and, on the other hand, the best choice of ρ also depends on different factors (e.g., the weight we wish to assign to the outliers in each specific example or the selection of the tuning parameters).

5 Concluding remarks

The Tukey biweight or bisquare family of loss functions has been used in order to compute fuzzy set-valued M-estimators of location through the Representer Theorem. The strong consistency and the robustness of this choice have been given. In future research, it would be interesting to develop a sensitivity analysis on how the selection of the involved tuning parameters affect the computation of M-estimators, as well as a deeper study of other families of loss functions for which the Representer Theorem still holds.

Acknowledgments

This research has been partially supported by the Spanish Ministry of Economy and Competitiveness through the Grant MTM2013-44212-P, the Principality of Asturias/FEDER Grant GRUPIN14-101, Grant C16/15/068 of International Funds KU Leuven and IAP research network grant nr. P7/06 of the Belgian government. Their support is gratefully acknowledged.

References

1. Beaton AE, Tukey JW (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16:147–185
2. Colubi A, González-Rodríguez G (2015) Fuzziness in data analysis: Towards accuracy and robustness. Fuzzy Sets Syst 281:260–271
3. Donoho DL, Huber PJ (1983) The notion of breakdown point. In: Bickel PJ, Doksum K, Hodges JL Jr (eds) A Festschrift for Erich L. Lehmann. Wadsworth, Belmont
4. F.R. Hampel, " A general qualitative definition of robustness," *Ann. Math. Stat.*, Vol. 42, no. 6, 1971, pp. 1887–1896.
5. Hampel FR (1974) The influence curve and its role in robust estimation. J Amer Stat Assoc 69:383–393
6. Hesketh T, Pryor R, Hesketh B (1988) An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. Int J ManMach Stud 29:21–35
7. Huber PJ (1964) Robust estimation of a location parameter. Ann Math Statist 35:73–101
8. Huber PJ (1981) Robust Statistics. J Wiley & Sons, New York
9. Kim JS, Scott CD (2012) Robust kernel density estimation. J Mach Learn Res 13:2529–2565
10. Puri ML, Ralescu DA (1986) Fuzzy random variables. J Math Anal Appl 114:409–422
11. Sinova B, Gil MA, Colubi A, Van Aelst S (2012) The median of a random fuzzy number. The 1-norm distance approach. Fuzzy Sets Syst 200:99–115
12. Sinova B, Gil MA, Van Aelst S. M-estimates of location for the robust central tendency of fuzzy data. IEEE Trans Fuzzy Syst. Accepted
13. Trutschnig W, González-Rodríguez G, Colubi A, Gil MA (2009) A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Inf Sci 179:3964–3972