On the consistency of a spatial-type interval-valued median for random intervals

Beatriz Sinova[a,c,*], Stefan Van Aelst[b,c]

[a]*Departamento de Estadística e I.O. y D.M.*
Universidad de Oviedo, 33071 Oviedo, Spain
[b]*Department of Mathematics, KU Leuven, 3001 Leuven, Belgium*
[c]*Department of Applied Mathematics, Computer Science and Statistics*
Ghent University, 9000 Gent, Belgium

Abstract

The d_θ-median is a much more robust estimator of the location of a random interval than the mean. We show that under general conditions the sample d_θ-median is a strongly consistent estimator of the d_θ-median.

Keywords: Interval-valued data, Random intervals, Spatial median, Consistency
2000 MSC: 62G35, 62–07

1. Introduction

In this data driven era, the amount and complexity of the available data grows at an almost incredible speed. Therefore, there is a high need to develop novel tools to cope with complex data structures, such as incomplete/missing data, functional data, interval valued or fuzzy data, and several other types of data.

Interval-valued data may arise for different reasons. The data may come from intrinsically interval-valued random elements (e.g. the daily fluctuation of the systolic blood pressure) or from random elements derived from an underlying real-valued random variable to preserve a level of confidentiality

*Corresponding author
Email addresses: sinovabeatriz@uniovi.es (Beatriz Sinova), stefan.vanaelst@wis.kuleuven.be (Stefan Van Aelst)

(e.g. when indicating an interval containing the real salary) or due to measurement imprecision (e.g. censored data), or from aggregation of a typically large dataset, etc. In this work no assumption about the source of the data is needed.

Interval-valued data are a type of complex data that requires specific statistical techniques. The main issue is that the space of intervals is only semilinear, but not linear due to the lack of the opposite of an interval. Therefore, although intervals can be identified with two-dimensional vectors (with first component the mid-point/centre and second component the non-negative spread/radius), it is not advisable to treat them as regular bivariate data. Indeed, common assumptions for multivariate techniques do not hold in this case.

Statistical procedures for random interval-valued data have already been proposed in the literature for different purposes, such as regression analysis (e.g. Gil *et al.*, 2002); testing hypotheses (e.g. González-Rodríguez *et al.*, 2012), clustering (e.g. D'Urso and Giordani *et al.*, 2006), principal component analysis (e.g. D'Urso and Giordani, 2004) and modelling distributions (see Brito and Duarte Silva, 2012).

A common location measure in this setting is the Aumann mean (Aumann, 1965). It is supported by several valuable properties and is also coherent with the interval arithmetic. The main disadvantage is that it is strongly influenced by outliers or small data changes, which can make it unreliable as a measure of the location of a random interval. In fact, it inherits this drawback from the standard real/vectorial-valued case.

In the real case, the most popular robust alternative is the median. For multivariate data the spatial median (or L_1-median, as introduced by Weber 1909) is a popular robust alternative. It is defined as the point in multivariate space with minimal average Euclidean distance to the observations. For more details and extensions, see for instance Gower (1974), Brown (1983), Milasevic and Ducharme (1987), Zuo (2013).

Sinova and Van Aelst (2014) adapted the spatial median to interval-valued data (see also Sinova *et al.* 2013), by using on this space as L^2 distance the versatile generalized metric introduced by Bertoluzza *et al.* (1995). The resulting d_θ-median estimator has been shown to be robust with high breakdown point and good finite-sample properties. In this paper we show another important property of the estimator, which is its strong consistency.

2. The d_θ-median of a random interval

Let $\mathcal{K}_c(\mathbb{R})$ denote the class of nonempty compact intervals. Any interval K in the space $K_c(\mathbb{R})$ can be characterized in terms of either its infimum and supremum, $K = [\inf K, \sup K]$, or its mid-point and spread or radius, $K = [\operatorname{mid} K - \operatorname{spr} K, \operatorname{mid} K + \operatorname{spr} K]$, where

$$\operatorname{mid} K = \frac{\inf K + \sup K}{2}, \quad \operatorname{spr} K = \frac{\sup K - \inf K}{2} \geq 0.$$

The usual interval arithmetic provides the addition, i.e. $K + K' = [\inf K + \inf K', \sup K + \sup K']$ with $K, K' \in \mathcal{K}_c(\mathbb{R})$ and the product by a scalar, i.e. $\gamma \cdot K = [\gamma \cdot \operatorname{mid} K - |\gamma| \cdot \operatorname{spr} K, \gamma \cdot \operatorname{mid} K + |\gamma| \cdot \operatorname{spr} K]$ with $K \in \mathcal{K}_c(\mathbb{R})$ and $\gamma \in \mathbb{R}$. With these two operations the space $\mathcal{K}_c(\mathbb{R})$ is only semilinear, so statistical techniques for interval-valued data will be based on distances.

The d_θ *metric* (Bertoluzza *et al.*, 1995, Gil *et al.*, 2002) can be defined as:

$$d_\theta(K, K') = \sqrt{(\operatorname{mid} K - \operatorname{mid} K')^2 + \theta \cdot (\operatorname{spr} K - \operatorname{spr} K')^2},$$

where $K, K' \in \mathcal{K}_c(\mathbb{R})$ and $\theta \in (0, \infty)$. A *random interval* can be defined as a Borel measurable mapping $X : \Omega \to \mathcal{K}_c(\mathbb{R})$, where $(\Omega, \mathcal{A}, P)$ is a probability space, with respect to $\mathcal{A}$ and the Borel σ-field generated by the topology induced on $\mathcal{K}_c(\mathbb{R})$ by the d_θ metric.

The well-known *Aumann mean value* is the interval, if it exists, given by

$$E[X] = [E(\operatorname{mid} X) - E(\operatorname{spr} X), E(\operatorname{mid} X) + E(\operatorname{spr} X)].$$

Moreover, it is the Fréchet expectation with respect to the d_θ metric, i.e., it is the unique interval that minimizes, over $K \in \mathcal{K}_c(\mathbb{R})$, the expression $E[(d_\theta(X, K))^2]$.

In Sinova and Van Aelst (2014) the d_θ*-median* of a random interval X is defined as the interval(s) $\mathrm{M}_\theta[X] \in \mathcal{K}_c(\mathbb{R})$ such that

$$E(d_\theta(X, \mathrm{M}_\theta[X])) = \min_{K \in \mathcal{K}_c(\mathbb{R})} E(d_\theta(X, K)),$$

whenever the involved expectations exist. Analogously, the sample d_θ-median statistic is defined as follows.

Let $(X_1, \ldots, X_n)$ be iid random intervals associated with a probability space $(\Omega, \mathcal{A}, P)$ and with realizations $\mathbf{x}_n = (x_1, \ldots, x_n)$. The *sample* d_θ*-median* (or medians) $\widehat{\mathrm{M}_\theta}[X]_n$ is (are) the random interval that for any $\mathbf{x}_n$ is (are) the solution(s) of the following optimization problem:

$$\min_{K\in\mathcal{K}_c(\mathbb{R})} \frac{1}{n}\sum_{i=1}^{n} d_\theta(x_i,K) = \min_{(y,z)\in\mathbb{R}\times[0,\infty)} \frac{1}{n}\sum_{i=1}^{n} \sqrt{(\operatorname{mid} x_i - y)^2 + \theta\cdot(\operatorname{spr} x_i - z)^2}$$

Sinova and Van Aelst (2014) showed the existence of the sample d_θ-median estimator and its uniqueness whenever not all the two-dimensional sample points $\{(\operatorname{mid} x_i, \operatorname{spr} x_i)\}_{i=1}^n$ are collinear. Moreover, the robustness was shown by its finite sample breakdown point (Donoho and Huber, 1983), which is given by $\frac{1}{n}\cdot\lfloor\frac{n+1}{2}\rfloor$, where $\lfloor\cdot\rfloor$ denotes the floor function.

3. Consistency of the sample d_θ-median

In this section we investigate the strong consistency of the sample d_θ-median under general conditions.

Theorem 1. *Let X be a random interval associated with a probability space $(\Omega,\mathcal{A},P)$ such that the d_θ-median exists and is unique. Then, the sample d_θ-median is a strongly consistent estimator of the d_θ-median, that is,*

$$\lim_{n\to\infty} d_\theta(\widehat{\mathrm{M}}_\theta[X]_n, M_\theta[X]) = 0 \quad a.s.[P].$$

Proof. Sufficient conditions for the strong consistency of an estimator are given in Huber (1967). We will check that these conditions, detailed below, are satisfied in our case and, hence, the theorem follows directly from Huber's general result.

First, note that the interval space ($\mathbb{R}\times[0,\infty)$ with the topology induced by the d_θ-metric) is a locally compact space with a countable base and $(\Omega,\mathcal{A},P)$ is a probability space.

Let $\rho(\omega,(y,z))$ be the following real-valued function on $\Omega\times(\mathbb{R}\times[0,\infty))$:

$$\begin{array}{rccl} \rho: & \Omega\times(\mathbb{R}\times[0,\infty)) & \longrightarrow & \mathbb{R} \\ & (\omega,(y,z)) & \longmapsto & d_\theta(X(\omega),[y-z,y+z]). \end{array}$$

Assuming that $\omega_1,\omega_2\ldots$ are independent Ω-valued random elements with common probability distribution P, the sequence of functions $\{T_n\}_{n\in\mathbb{N}}$, defined as $T_n(\omega_1,\ldots,\omega_n) = \widehat{\mathrm{M}}_\theta[(X(\omega_1),\ldots,X(\omega_n))]_n$, satisfies that

$$\frac{1}{n}\sum_{i=1}^{n} d_\theta(X(\omega_i),T_n(\omega_1,\ldots,\omega_n)) - \inf_{(y,z)\in\mathbb{R}\times[0,\infty)} \frac{1}{n}\sum_{i=1}^{n} d_\theta(X(\omega_i),[y-z,y+z]) \underset{n\to\infty}{\longrightarrow} 0$$

almost surely (obviously because of the definition of the sample d_θ-median).

We now recall the sufficient conditions for the strong consistency in Huber (1967).

Condition (A-1). For each fixed $(y_0, z_0) \in \mathbb{R} \times [0, \infty)$, the function

$$\begin{array}{rcll} \rho_0 : \ \Omega & \longrightarrow & \mathbb{R} & \\ \omega & \longmapsto & \rho(\omega, (y_0, z_0)) & = d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) \end{array}$$

is $\mathcal{A}$-measurable and separable in Doob's sense, i.e. there is a P-null set N and a countable subset $S \subset \mathbb{R} \times [0, \infty)$ such that for every open set $U \subset \mathbb{R} \times [0, \infty)$ and every closed interval A, the sets

$$V_1 = \{\omega : \rho(\omega, (y, z)) \in A, \forall (y, z) \in U\}$$

$$V_2 = \{\omega : \rho(\omega, (y, z)) \in A, \forall (y, z) \in U \cap S\}$$

differ by at most a subset of N.

Condition (A-2). The function ρ is a.s. lower semicontinuous in (y_0, z_0), i.e.

$$\inf_{(y,z) \in U} \rho(\omega, (y, z)) \longrightarrow \rho(\omega, (y_0, z_0)), \tag{1}$$

as the neighborhood U of (y_0, z_0) shrinks to $\{(y_0, z_0)\}$.

Condition (A-3). There is a measurable function $a : \Omega \to \mathbb{R}$ such that

$$E[\rho(\omega, (y, z)) - a(\omega)]^- < \infty \quad \text{ for all } (y, z) \in \mathbb{R} \times [0, \infty),$$

$$E[\rho(\omega, (y, z)) - a(\omega)]^+ < \infty \quad \text{ for some } (y, z) \in \mathbb{R} \times [0, \infty).$$

Thus, $\gamma((y, z)) = E[\rho(\omega, (y, z)) - a(\omega)]$ is well-defined for all (y, z).

Condition (A-4). There is a $(y_0, z_0) \in \mathbb{R} \times [0, \infty)$ such that $\gamma((y, z)) > \gamma((y_0, z_0))$ for all $(y, z) \neq (y_0, z_0)$.

Condition (A-5). There is a continuous function $b((y, z)) > 0$ such that

- for some integrable h, $\inf_{(y,z) \in \mathbb{R} \times [0,\infty)} \frac{\rho(\omega,(w,z)) - a(\omega)}{b((y,z))} \geq h(\omega)$.
- $\liminf_{(y,z) \to \infty} b((y, z)) > \gamma((y_0, z_0))$.
- $E\left[\liminf_{(y,z) \to \infty} \frac{\rho(\omega,(y,z)) - a(\omega)}{b((y,z))}\right] \geq 1$.

We now verify these conditions.

(A-1) For each fixed $(y_0, z_0) \in \mathbb{R} \times [0, \infty)$, the function ρ_0 is $\mathcal{A}$-measurable because $\operatorname{mid} X$ and $\operatorname{spr} X$ are measurable functions since X is a random interval. ρ_0 is also separable in Doob's sense: choose $S = \mathbb{Q} \times (\mathbb{Q} \cap [0, \infty))$ as countable subset, then for every open set $U \subset \mathbb{R} \times [0, \infty)$ and every closed interval A, it can be seen that the following sets coincide.

$$V_1 = \{\omega : \rho_0(\omega) \in A, \forall (y, z) \in U\}, \ V_2 = \{\omega : \rho_0(\omega) \in A, \forall (y, z) \in U \cap S\}$$

Obviously, $V_1 \subseteq V_2$. By *reductio ad absurdum*, suppose that $V_2 \cap V_1^c \neq \emptyset$. Let $\omega_0 \in V_2 \cap V_1^c$:

- Since $\omega_0 \in V_2$, $\rho(\omega_0, (y, z)) \in A$ for all $(y, z) \in U \cap S$;
- Since $\omega_0 \in V_1^c$, there exists $(y_0, z_0) \in U$ such that $\rho(\omega_0, (y_0, z_0)) \in A^c$. A^c is an open set, so there exists a ball of radius $r > 0$ such that

$$(\rho(\omega_0, (y_0, z_0)) - r, \rho(\omega_0, (y_0, z_0)) + r) \subseteq A^c.$$

Now, notice that for a fixed $\omega \in \Omega$, the following function is continuous:

$$\begin{array}{rccc} \rho_\omega : & \mathbb{R} \times [0, \infty) & \longrightarrow & \mathbb{R} \\ & (y, z) & \longmapsto & \rho(\omega, (y, z)). \end{array}$$

Thus, $B = \rho_{\omega_0}^{-1}(\rho(\omega_0, (y_0, z_0)) - r, \rho(\omega_0, (y_0, z_0)) + r)$ is an open set of $\mathbb{R} \times [0, \infty)$ and $U \cap B \neq \emptyset$. S is a dense set of $\mathbb{R} \times [0, \infty)$, so $U \cap B \cap S \neq \emptyset$. Let $(y', z') \in U \cap B \cap S$. Then, $(y', z') \in U \cap S$, so $\rho(\omega_0, (y', z')) \in A$. But also,

$$\rho(\omega_0, (y', z')) \in (\rho(\omega_0, (y_0, z_0)) - r, \rho(\omega_0, (y_0, z_0)) + r) \subset A^c.$$

This is a contradiction, so the conclusion is that also $V_2 \subseteq V_1$, which proofs this condition.

(A-2) Let ω be any element of Ω and let (y_0, z_0) be any (fixed) point of $\mathbb{R} \times [0, \infty)$. First, note that (1) is fulfilled for a sequence of neighborhoods $\{U_n\}_{n \in \mathbb{N}}$ of (y_0, z_0) with $U_n \supseteq U_{n+1}$ for all n. Since

$$\left\{ \inf_{(y,z) \in U_n} d_\theta(X(\omega), [y - z, y + z]) \right\}_{n \in \mathbb{N}}$$

is a monotonically increasing sequence and $(y_0, z_0) \in \cap_{n \in \mathbb{N}} U_n$, the sequence is bounded by $d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0])$ and, thus, it converges to its supremum. We now show that this supremum is indeed $d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0])$.

By *reductio ad absurdum*, suppose there is a smaller upper bound $c = d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) - \varepsilon$, for an arbitrary $\varepsilon > 0$. Let U_{n_0} denote a neighborhood of (y_0, z_0) such that $U_{n_0} \subseteq C = B((y_0, z_0), \frac{\varepsilon}{2})$. It follows that $c < \inf_{(y,z)\in U_{n_0}} d_\theta(X(\omega), [y - z, y + z])$, so c cannot be the supremum. Indeed,

$$\inf_{(y,z)\in U_{n_0}} d_\theta(X(\omega), [y - z, y + z]) \geq \inf_{(y,z)\in C} d_\theta(X(\omega), [y - z, y + z])$$

$$\geq \inf_{(y,z)\in C} [d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) - d_\theta([y - z, y + z], [y_0 - z_0, y_0 + z_0])]$$

$$= d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) - \sup_{(y,z)\in C} d_\theta([y - z, y + z], [y_0 - z_0, y_0 + z_0])$$

$$> d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) - \varepsilon = c.$$

Now, we extend this result to general sequences $\{U_n\}_{n\in\mathbb{N}}$. Consider the suprema and the infima radii reached in every neighborhood, namely,

$$r_n = \sup_{(y,z)\in U_n} d_\theta([y_0 - z_0, y_0 + z_0], [y - z, y + z]),$$

$$s_n = \inf_{(y,z)\in U_n} d_\theta([y_0 - z_0, y_0 + z_0], [y - z, y + z]).$$

It follows that $r_n \underset{n\to\infty}{\longrightarrow} 0$, since $\{U_n\}_{n\in\mathbb{N}}$ shrinks to $\{(y_0, z_0)\}$. Hence, also $s_n \underset{n\to\infty}{\longrightarrow} 0$ since $0 \leq s_n \leq r_n$ for all $n \in \mathbb{N}$.

Choose $\varepsilon > 0$ arbitrarily. As $r_n \underset{n\to\infty}{\longrightarrow} 0$, there exists $n_1 \in \mathbb{N}$ such that for all $n > n_1$, $r_n < \varepsilon$. Then, $U_n \subseteq B((y_0, z_0), r_n)$ and

$$\inf_{(y,z)\in U_n} d_\theta(X(\omega), [y - z, y + z]) \geq \inf_{(y,z)\in B((y_0,z_0),r_n)} d_\theta(X(\omega), [y - z, y + z])$$

$$\geq d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) - \sup_{(y,z)\in B((y_0,z_0),r_n)} d_\theta([y_0 - z_0, y_0 + z_0], [y - z, y + z])$$

$$> d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) - \varepsilon.$$

Analogously, as $s_n \underset{n\to\infty}{\longrightarrow} 0$, there exists $n_2 \in \mathbb{N}$ such that for all $n > n_2$, $s_n < \varepsilon$. Therefore, $U_n \supseteq B((y_0, z_0), s_n)$ and

$$\inf_{(y,z)\in U_n} d_\theta(X(\omega), [y - z, y + z]) \leq \inf_{(y,z)\in B((y_0,z_0),s_n)} d_\theta(X(\omega), [y - z, y + z])$$

$$\leq d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) + \inf_{(y,z)\in B((y_0,z_0),s_n)} d_\theta([y - z, y + z], [y_0 - z_0, y_0 + z_0])$$

$$< d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) + \varepsilon.$$

So for any $\varepsilon > 0$, there exists $n_0 = \max\{n_1, n_2\}$ s.t. for all $n > n_0$, $\left| \inf_{(y,z)\in U_n} d_\theta(X(\omega), [y-z, y+z]) - d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0]) \right| < \varepsilon$. That is, the considered sequence converges to $d_\theta(X(\omega), [y_0 - z_0, y_0 + z_0])$, which completes the proof of (1).

(A-3) Let a be the measurable function (see (A-1)):

$$\begin{array}{rccl} a: & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & d_\theta(X(\omega), [0,0]) = \sqrt{(\operatorname{mid} X(\omega))^2 + \theta \cdot (\operatorname{spr} X(\omega))^2}. \end{array}$$

For any $(y, z) \in \mathbb{R} \times [0, \infty)$, we then have that $E[\rho(\omega, (y, z)) - a(\omega)]^-$

$$= \int_\Omega - \min\{d_\theta(X(\omega), [y-z, y+z]) - d_\theta(X(\omega), [0,0]), 0\}\, dP(\omega)$$

$$= \int_{\substack{\{\omega\in\Omega\,:\, d_\theta(X(\omega),[0,0]) \\ > d_\theta(X(\omega),[y-z,y+z])\}}} \big[d_\theta(X(\omega), [0,0]) - d_\theta(X(\omega), [y-z, y+z])\big]\, dP(\omega).$$

By the triangular inequality,

$$\leq \int_{\substack{\{\omega\in\Omega\,:\, d_\theta(X(\omega),[0,0]) \\ > d_\theta(X(\omega),[y-z,y+z])\}}} \big[d_\theta(X(\omega), [y-z, y+z]) + d_\theta([y-z, y+z], [0,0])$$

$$-d_\theta(X(\omega), [y-z, y+z])\big]\, dP(\omega)$$

$$= d_\theta([y-z, y+z], [0,0]) \cdot P\big(\omega : d_\theta(X(\omega), [0,0]) > d_\theta(X(\omega), [y-z, y+z])\big) < \infty.$$

Analogously, $E[\rho(\omega, (y, z)) - a(\omega)]^+$

$$= \int_\Omega \max\{d_\theta(X(\omega), [y-z, y+z]) - d_\theta(X(\omega), [0,0]), 0\}\, dP(\omega)$$

$$\leq d_\theta([0,0], [y-z, y+z]) \cdot P\big(\omega : d_\theta(X(\omega), [0,0]) \leq d_\theta(X(\omega), [y-z, y+z])\big) < \infty,$$

for all $(y, z) \in \mathbb{R} \times [0, \infty)$. So, with this choice for the function a, both inequalities in condition *(A-3)* hold.

(A-4) The d_θ-median exists and is unique, so that

$$(\operatorname{mid} M_\theta[X], \operatorname{spr} M_\theta[X]) = \arg \min_{(y,z)\in\mathbb{R}\times[0,\infty)} E\left[d_\theta(X(\omega), [y-z, y+z])\right]$$

$$= \arg \min_{(y,z)\in\mathbb{R}\times[0,\infty)} E\left[d_\theta(X(\omega), [y-z, y+z])\right] - E\left[d_\theta(X(\omega), [0,0])\right]$$

$$= \arg \min_{(y,z)\in\mathbb{R}\times[0,\infty)} \gamma((y,z)).$$

Thus, $(y_0, z_0) := (\operatorname{mid} M_\theta[X], \operatorname{spr} M_\theta[X])$ fulfills this condition.

(A-5) Consider the continuous function $b((y,z)) > 0$ defined as

$$\begin{array}{rccl} b: & \mathbb{R}\times[0,\infty) & \longrightarrow & \mathbb{R} \\ & (y,z) & \longmapsto & d_\theta([y-z,y+z],[0,0])+1. \end{array}$$

- For the integrable function $h(\omega) := -1$, it then holds that

$$\inf_{(y,z)\in\mathbb{R}\times[0,\infty)} \frac{d_\theta(X(\omega),[y-z,y+z]) - d_\theta(X(\omega),[0,0])}{d_\theta([y-z,y+z],[0,0])+1} \geq -1$$

  because using the triangular inequality,

$$\geq \inf_{(y,z)\in\mathbb{R}\times[0,\infty)} \frac{d_\theta(X(\omega),[0,0]) - d_\theta([y-z,y+z],[0,0]) - d_\theta(X(\omega),[0,0])}{d_\theta([y-z,y+z],[0,0])+1}$$

$$= \inf_{(y,z)\in\mathbb{R}\times[0,\infty)} \frac{-d_\theta([y-z,y+z],[0,0])}{d_\theta([y-z,y+z],[0,0])+1} \geq -1.$$

- The condition $\liminf_{(y,z)\to\infty} b((y,z)) > \gamma((y_0,z_0))$ is satisfied. Let $\{(y_n,z_n)\} \subset \mathbb{R}\times[0,\infty)$ be any sequence with $(y_n,z_n) \underset{n\to\infty}{\longrightarrow} \infty$ in the sense that $d_\theta([y_n-z_n,y_n+z_n],[0,0]) \underset{n\to\infty}{\longrightarrow} \infty$, and set

  $M = E\left[d_\theta(X(\omega),[y_0-z_0,y_0+z_0]) - d_\theta(X(\omega),[0,0])\right] = \gamma((y_0,z_0)) \in \mathbb{R}$,

  where (y_0,z_0) represents the minimum in (A-4). Then, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $d_\theta([y_n-z_n,y_n+z_n],[0,0]) > M$. Thus,

$$\inf_{k\geq n_0} b((y_k,z_k)) = \inf_{k\geq n_0} \left(d_\theta([y_k-z_k,y_k+z_k],[0,0])+1\right) \geq M+1,$$

  so $\liminf_{n\to\infty} b((y_n,z_n)) = \lim_{n\to\infty}\left(\inf_{k\geq n} b((y_k,z_k))\right) \geq M+1 > M = \gamma((y_0,z_0))$.

- It also holds that

$$E\left[\liminf_{(y,z)\to\infty} \frac{d_\theta(X(\omega),[y-z,y+z]) - d_\theta(X(\omega),[0,0])}{b((y,z))}\right] \geq 1, \qquad (2)$$

  because we have that

$$\liminf_{(y,z)\to\infty} \frac{d_\theta(X(\omega),[y-z,y+z]) - d_\theta(X(\omega),[0,0])}{d_\theta([y-z,y+z],[0,0])+1} \geq 1,$$

Indeed, for any $\omega \in \Omega$

$$\liminf_{(y,z)\to\infty} \frac{d_\theta(X(\omega), [y-z, y+z]) - d_\theta(X(\omega), [0,0])}{d_\theta([y-z, y+z], [0,0]) + 1}$$

$$= \lim_{n\to\infty} \left(\inf_{k\geq n} \frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \right).$$

The sequence

$$\left\{ \inf_{k\geq n} \frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \right\}_{n\in\mathbb{N}} \tag{3}$$

is monotonically increasing with upper bound 1, since for all $k \in \mathbb{N}$, using the triangular inequality,

$$\frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1}$$

$$\leq \frac{d_\theta([y_k - z_k, y_k + z_k], [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \leq 1.$$

Hence, the sequence (3) converges to its supremum:

$$\lim_{n\to\infty} \left(\inf_{k\geq n} \frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \right)$$

$$= \sup_n \left(\inf_{k\geq n} \frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \right)$$

Finally, it will be checked this supremum is at least equal to 1. By *reductio ad absurdum*, we suppose that

$$\sup_n \left(\inf_{k\geq n} \frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \right) = 1 - \varepsilon,$$

for some $\varepsilon > 0$. We need to show that we can find an $n^* \in \mathbb{N}$ such that

$$\inf_{k\geq n^*} \frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} > 1 - \varepsilon. \tag{4}$$

Recall that $(y_n, z_n) \underset{n\to}{\longrightarrow} \infty$, so for any $M \in \mathbb{R}$, there exists $n^* \in \mathbb{N}$ such that for all $n \geq n^*$, $d_\theta([y_n - z_n, y_n + z_n], [0,0]) > M$. Then,

$$d_\theta([y_n - z_n, y_n + z_n], X(\omega)) \geq d_\theta([y_n - z_n, y_n + z_n], [0,0]) - d_\theta(X(\omega), [0,0])$$

$$> M - d_\theta(X(\omega), [0,0]).$$

Take $M := \frac{2}{\varepsilon} - 1 + \frac{4}{\varepsilon} \cdot d_\theta(X(\omega), [0,0]) \in \mathbb{R}$ (for the fixed arbitrary $\omega \in \Omega$), then for any $k \geq n^*$, we have that

$$\begin{aligned}
&d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0]) \\
&= \left(1 - \frac{\varepsilon}{2}\right) d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) + \frac{\varepsilon}{2} d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) \\
&\quad - d_\theta(X(\omega), [0,0]) \\
&\geq \left(1 - \frac{\varepsilon}{2}\right) d_\theta([y_k - z_k, y_k + z_k], [0,0]) - \left(1 - \frac{\varepsilon}{2}\right) d_\theta(X(\omega), [0,0]) \\
&\quad + \frac{\varepsilon}{2} d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0]) \\
&> \left(1 - \frac{\varepsilon}{2}\right) d_\theta([y_k - z_k, y_k + z_k], [0,0]) + \frac{\varepsilon}{2}(M - d_\theta(X(\omega), [0,0])) \\
&\quad - \left(2 - \frac{\varepsilon}{2}\right) d_\theta(X(\omega), [0,0]) \\
&> \left(1 - \frac{\varepsilon}{2}\right) d_\theta([y_k - z_k, y_k + z_k], [0,0]) + \frac{\varepsilon}{2} \left(\frac{2}{\varepsilon} - 1 + \left(\frac{4}{\varepsilon} - 1\right) d_\theta(X(\omega), [0,0])\right) \\
&\quad - \left(2 - \frac{\varepsilon}{2}\right) d_\theta(X(\omega), [0,0]) = \left(1 - \frac{\varepsilon}{2}\right) d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1 - \frac{\varepsilon}{2} \\
&= \left(1 - \frac{\varepsilon}{2}\right) \left(d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1\right). \qquad \square
\end{aligned}$$

Hence, for all $k \geq n^*$,

$$\frac{d_\theta(X(\omega), [y_k - z_k, y_k + z_k]) - d_\theta(X(\omega), [0,0])}{d_\theta([y_k - z_k, y_k + z_k], [0,0]) + 1} \geq 1 - \frac{\varepsilon}{2} > 1 - \varepsilon,$$

which implies (4) and thus the inequality (2) follows.Hence, the three inequalities in condition *(A-5)* are satisfied, which completes the proof.

4. Concluding remarks

This paper complements the study of the properties of the d_θ-median as a robust estimator of the center of a random interval by showing its strong consistency, which is one of the most important basic properties of an estimator. This result open the door to further develop robust statistical inference for random intervals based on the d_θ-median.

5. Acknowledgements

Authors are grateful to María Ángeles Gil for her helpful suggestions to improve this paper. The research by Sinova was partially supported by/benefited from the Spanish Ministry of Science and Innovation Grant MTM2009-09440-C02-01, the Ayuda del Programa de FPU AP2009-1197 from the Spanish Ministry of Education and the Ayuda para Estancias Breves del Programa FPU EST12/00344. The research by Van Aelst was supported by a grant of the Fund for Scientific Research-Flanders (FWO-Vlaanderen) and by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). Their financial support is gratefully acknowledged.

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

[2] Bertoluzza, C., Corral, N., Salas, A., 1995. On a new class of distances between fuzzy numbers. Math. & Soft Comput. 2, 71–84.

[3] Brito, P., Duarte Silva, A.P., 2012. Modelling interval data with Normal and Skew-Normal distributions. J. Appl. Stat. 39(1), 3–20.

[4] Brown, B.M., 1983. Statistical uses of the spatial median. J. Royal Stat. Soc. Ser. B 45(1), 25–30.

[5] Donoho, D.L., Huber, P.J., 1983. The notion of breakdown point, in: Bickel, P.J., Doksum, K., Hodges Jr, J.L. (Eds.), A Festschrift for Erich L. Lehmann. Wadsworth, Belmont.

[6] D'Urso, P., Giordani, P., 2004. A least squares approach to principal component analysis for interval valued data. Chemometr. Intell. Lab. 70(2), 179–192.

[7] D'Urso, P., Giordani, P., 2006. A robust fuzzy k-means clustering model for interval valued data. Comput. Stat. 21(2), 251–269.

[8] Fagundes, R.A.A., De Souza, R.M.C.R., Cysneiros, F.J.A., 2013. Robust regression with application to symbolic interval data. Eng. Appl. Art. Intel. 26(1), 564–573.

[9] Gil, M.A., Lubiano, M.A., Montenegro, M., López-García, M.T., 2002. Least squares fitting of an affine function and strength of association for interval data. Metrika 56, 97–111.

[10] González-Rodríguez, G., Colubi, A., Gil, M.A., 2012. Fuzzy data treated as functional data: A one-way ANOVA test approach. Comput. Stat. Data Anal. 56, 943–955.

[11] Gower, J.C., 1974. Algorithm AS 78: The mediancentre. Appl. Statist. 23, 466–470.

[12] Huber, P.J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, pp. 221–233.

[13] Milasevic, P., Ducharme, G.R., 1987. Uniqueness of the spatial median. Ann. Statist. 15, 1332–1333.

[14] Sinova, B., González-Rodríguez, G., Van Aelst, S., 2013. An alternative approach to the median of a random interval using an L^2 metric, in: Kruse, R., Berthold, M.R., Moewes, C., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (Eds.), Sinergies of Soft Computing and Statistics for Intelligent Data Analysis. Springer, Heidelberg, pp. 273–281.

[15] Sinova, B., Van Aelst, S., 2014. A spatial-type interval-valued median for random intervals. Submitted

[16] Weber, A., 1909. Über den Standort der Industrien. Mohr, Tübingen.

[17] Zuo, Y., 2013. Multidimensional medians and uniqueness. Comput. Stat. Data Anal. 66, 82–88.