On some concepts related to star-shaped sets

Ana Belén Ramos-Guajardo[1], Gil González-Rodríguez[1], Ana Colubi[1], Maria Brigida Ferraro[2], and Ángela Blanco-Fernández[1]

Abstract The convenient theoretical properties of the support function and the Minkowski addition-based arithmetic have been shown to be useful when dealing with compact and convex sets on $\mathbb{R}^p$. However, both concepts present several drawbacks in certain contexts. The use of the radial function instead of the support function is suggested as an alternative to characterize a wider class of sets - the so-called star-shaped sets - which contains the class of compact and convex sets as a particular case. The concept of random star-shaped set is considered, and some statistics for this kind of variable are shown. Finally, some measures for comparing star-shaped sets are introduced.

1 Introduction

Random sets, also called set-valued random variables and denoted by RSs for short, have been used in different fields. For instance, they have been shown to be useful in spatial data analysis [18], in Econometrics [3] and in Structural Engineering [25], to name but a few. RSs can also be viewed as imprecise random variables, as Pedro Gil and his colleagues have pointed out in [17]. Several results for RSs have been accomplished, such as limit theorems [1, 19], confidence sets for the (Aumann) expected value [4], hypothesis testing for the expected value or the (Fréchet) variance [11, 13, 14, 15, 20, 21] and inference on regression models [2, 9, 12].

In the one-dimensional case, the compact intervals A of $\mathbb{R}$ can be characterized by either the infima and suprema of A, $(\inf A, \sup A)$ so that $\inf A < \sup A$, or by the mid-point and radius of A, $(\operatorname{mid} A, \operatorname{spr} A) \in \mathbb{R} \times \mathbb{R}^+$. The usual interval arithmetic is based on the Minkowski addition [16] and

INDUROT/Dept. of Statistics, OR and MD, University of Oviedo, Spain {blancoangela,colubi,gil,ramosana}@uniovi.es · Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Italy mariabrigida.ferraro@uniroma1.it

the product by a scalar, and it preserves the length of the resulting intervals. Many statistical results concerning interval data are based on the Minkowski arithmetic (see, for instance, [2, 7, 9, 12, 15, 21]).

The statistical studies developed until now for the p-dimensional situation (with $p > 1$) frequently take advantage of some convenient theoretical properties of the support function, but they have several drawbacks in certain situations. To overcome these drawbacks, an alternative to the support function for characterizing star-shaped sets by means of the so-called *radial/polar function* has been investigated in [5, 6].

Basic concepts related to this new representation and some statistical results are addressed. More concretely, the concept of random star-shaped set - i.e. a random variable taking star-shaped sets as outcomes -, and those of expected value and variance are considered. The corresponding sample moments are defined, and the consistency with respect to their population counterparts is highlighted. In addition, the concept of mean directional length is introduced and some comparative measures of centered star-shaped sets are suggested.

The rest of the paper is organized as follows. Section 2 is devoted to the introduction of some preliminaries regarding compact and convex sets and star-shaped sets. The concept of random star-shaped sets and their moments are recalled in Section 3. A basic example illustrating these sample moments is provided. The notions related to the mean directional length are discussed in Section 4. Finally, some conclusions and open problems are provided in Section 5.

2 Preliminaries

Let the space $\mathbb{R}^p$ be endowed with the Euclidean norm $\|\cdot\|$ and the corresponding inner product $\langle \cdot, \cdot \rangle$. Let $\mathbb{S}^{p-1} = \{u \in \mathbb{R}^p : \|u\| = 1\}$ be the hypersphere with radius 1. The space of all non-empty compact and convex subsets of $\mathbb{R}^p$ is denoted by $\mathcal{K}_c(\mathbb{R}^p)$. If $A \in \mathcal{K}_c(\mathbb{R}^p)$, then the *support function* of U is defined such that $s_A(u) = \sup_{a \in A} \langle u, a \rangle$ for $u \in \mathbb{S}^{p-1}$ [10, 19].

The location and the imprecision of a set $A \in \mathcal{K}_c(\mathbb{R}^p)$ can be determined in terms of the support function by the so-called mid-spread representation in such a way that $s_A = \text{mid}_A + \text{spr}_A$, where $\text{mid}_A(u) = (s_A(u) - s_A(-u))/2$ and $\text{spr}_A(u) = (s_A(u) + s_A(-u))/2$ for all $u \in \mathbb{S}^{p-1}$.

As shown in Figure 1, the support function identifies the boundary of the corresponding set, but the obtained result is not easy to relate with the original shape of the set. Actually, it is very difficult to identify which is the original set associated with a function verifying the properties of the support function (if any). This could be a drawback in some applied problems in

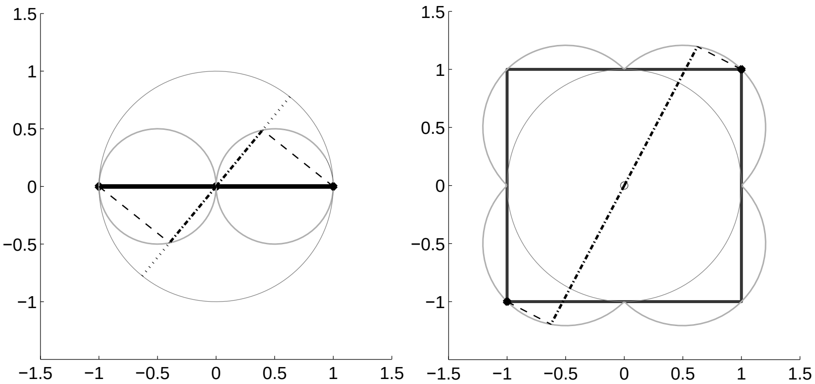

Fig. 1 Support function (distance from (0,0) to the contour of the gray line, marked in dashed-dotted black) of a line and a square in $\mathbb{R}^2$ (in black)

which it is necessary to clearly identify the shape of the sets (as, for instance, in image processing).

To overcome this disadvantage, other characterizations of sets can be taken into account. For instance, a useful tool in this framework is the so-called *radial function* [24]. It is defined on the class of *star-shaped sets* of $\mathbb{R}^p$, denoted by $\mathcal{K}_s(\mathbb{R}^p)$, which is an extension of $\mathcal{K}_c(\mathbb{R}^p)$ - i.e. $\mathcal{K}_c(\mathbb{R}^p) \subset \mathcal{K}_s(\mathbb{R}^p)$. A star-shaped set $A \in \mathcal{K}_s(\mathbb{R}^p)$ with respect to k_A, where k_A is *a center* of A, is a nonempty compact subset of $\mathbb{R}^p$ such that for all $a \in A$, $\lambda k_A + (1-\lambda)a \in A$ for all $\lambda \in [0,1]$. The radial function of a star-shaped set A is defined as $\rho_A : \mathbb{S}^{p-1} \to \mathbb{R}^+$ so that $\rho_A(u) = \sup\{\lambda \geq 0 : k_A + \lambda u \in A\}$. In this context, k_A can be viewed as a location point of the star-shaped set A whereas ρ_A is related to the imprecision of the set. The formal definition of k_A and ρ_A to be used in statistical problems is not trivial. This problem has been recently addressed in [6]. From now on, this representation of sets will be called *center-radial characterization*.

In contrast to the support function, the radial function identifies the shape of the sets in an intuitive way, as it is shown in Figure 2, because it is simply based on the well-known polar coordinates over the unit sphere. In the case of the line (left side image in Figure 2), the radial function is equal to 0 for all $u \in \mathbb{S}$ except for $u_1 = (1,0)$ and $u_2 = (-1,0)$, with $\rho_A(u_1) = \rho_A(u_2) = 1$. Further advantages of the radial function with respect the support function are pointed out in [6].

The space $\mathcal{K}_s(\mathbb{R}^p)$ can be embedded into a cone on the Hilbert space $\mathcal{H}_r = \mathbb{R}^p \times \mathcal{L}^2(\mathbb{S}^{p-1})$ through the center-radial characterization. For the theoretical developments, from now on, star-shaped sets in $\mathcal{K}_s^*(\mathbb{R}^p)$ will be considered, where

$$\mathcal{K}_s^*(\mathbb{R}^p) = \left\{A \in \mathcal{K}_s(\mathbb{R}^p) | \rho_A \in \mathcal{L}^2(\mathbb{S}^{p-1})\right\}. \tag{1}$$

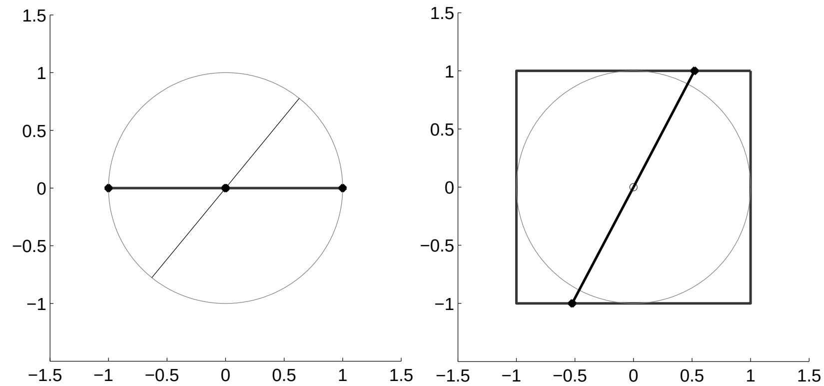

Fig. 2 Radial function of a line (black dots) and a square (which corresponds exactly to the square) in $\mathbb{R}^2$

Regarding the arithmetic, we could consider the Minkowski addition between two star-shaped sets A and B, $A + B = \{a + b \,|\, a \in A, b \in B\}$. However, it has been shown that the Minkowski addition is not always meaningful (see [5, 18]), and it does not agrees with the natural arithmetic induced by the center-radial characterization from the Hilbert space. That is, $A +_r \lambda B$ should be the element in $\mathcal{K}_s^*(\mathbb{R}^p)$ satisfying that $k_{A+_r\lambda B} = k_A + \lambda k_B$ and $\rho_{A+_r\lambda B} = \rho_A + \lambda\rho_B$, where $+$ denotes the usual sum of two points in R^p and the usual sum of two functions in $\mathcal{L}^2(\mathbb{S}^{p-1})$. An example of the differences between the Minkowski sum and the center-radial sum of two star-shaped sets in $\mathcal{K}_s^*(\mathbb{R}^p)$ is provided in Figure 3. We observe that the center-radial sum preserves, *directionally*, the lengths, whereas the Minkowski sum *dilates* them.

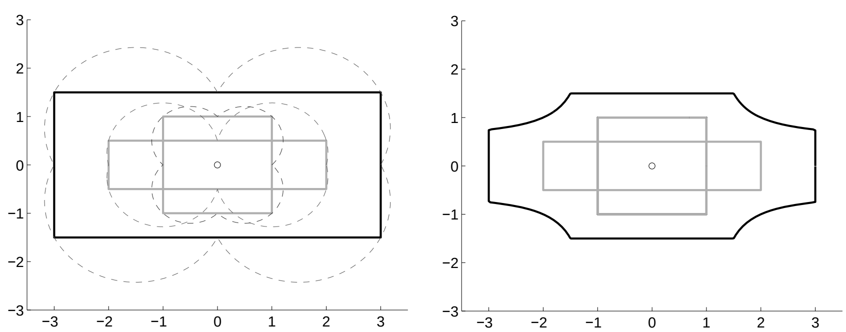

Fig. 3 Minkowski (left) and radial (right) sums (in black) of the gray quadrilaterals

Regarding the metric structure in $\mathcal{K}_s^*(\mathbb{R}^p)$, the center-radial characterization induces a natural family of distances from the corresponding one in the associated Hilbert space. Thus, for any two star-shaped sets $A, B \in \mathcal{K}_s^*(\mathbb{R}^p)$, the τ-metric is defined as

$$d_\tau(A,B) = \sqrt{\tau\|k_A - k_B\|^2 + (1-\tau)\|\rho_A - \rho_B\|_p^2}, \tag{2}$$

where $\tau \in (0,1)$ determines the importance given to the location in contrast to the imprecision, $\|\cdot\|$ denotes the usual norm in $\mathbb{R}^p$ and $\|\cdot\|_p$ is the usual L^2-type norm in $\mathcal{L}^2(\mathbb{S}^{p-1})$ [6].

3 Random star-shaped sets

Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathbb{R}^p \times \mathcal{K}_s^*(\mathbb{R}^p)$ is a *random star-shaped set* if it is a Borel measurable mapping with respect to $\mathcal{A}$ and the Borel σ-field generated by the topology induced by the metric d_τ on $\mathbb{R}^p \times \mathcal{K}_s^*(\mathbb{R}^p)$. Equivalently, X can be decomposed in terms of its center-radial characterization, that is, $X = (k_X, \rho_X)$, and X can be defined to be a random star-shaped set iff k_X and ρ_X are random elements in the real and functional framework respectively [6].

Now we are in a position to define some summarizing measures for random star-shaped sets. On one hand, if $E(\|k_X\|) < \infty$ and $E(\|\rho_X\|_p) < \infty$, then the *expected value* of X is defined as the element $E(X) \in \mathbb{R}^p \times \mathcal{K}_s^*(\mathbb{R}^p)$ so that $k_{E(X)} = E(k_X)$ and $\rho_{E(X)} = E(\rho_X)$ - this last expectation being considered in terms of the Bochner integral in $\mathcal{L}^2(\mathbb{S}^{p-1})$.

From an empirical point of view, given X a random star-shaped set and $\{X_i\}_{i=1}^n$ an i.i.d. sequence of random star-shaped sets drawn from X, the sample expectation of X can be defined in terms of the arithmetic in $\mathbb{R}^p \times \mathcal{K}_s^*(\mathbb{R}^p)$ as follows:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{3}$$

It is easy to show that $(k_{\overline{X}}, \rho_{\overline{X}}) = (\overline{k_X}, \overline{\rho_X})$.

If $E(\|k_X\|^2) < \infty$ and $E(\|\rho_X\|_p^2) < \infty$, then $E(X)$ is the unique element in $\mathbb{R}^p \times \mathcal{K}_s^*(\mathbb{R}^p)$ satisfying that

$$E(d_\tau^2(X, E(X))) = \min_{(k,A)\in\mathbb{R}^p\times\mathcal{K}_s^*(\mathbb{R}^p)} E(d_\tau^2(X,(k,A))). \tag{4}$$

Thus, by following the Fréchet approach, the *(scalar) variance* of a random star-shaped set X, denoted by σ_X^2, is defined as

$$\sigma_X^2 = E(d_\tau^2(X, E(X))). \tag{5}$$

The sample variance is also defined in terms of the distance d_τ, or equivalently, in terms of the corresponding variances in $\mathbb{R}^p$ and $\mathcal{L}^2(\mathbb{S}^{p-1})$, as follows:

$$\widehat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} d_\tau^2(X_i, \overline{X}) = (1-\tau)\widehat{\sigma}_{k_X}^2 + \tau\widehat{\sigma}_{\rho_X}^2 \,. \tag{6}$$

The consistency of the estimators (3) and (5) for the mean and the variance of random star-shaped sets, respectively, is provided in the following result. It is an immediate consequence of the Strong Law for Large Numbers in Banach spaces.

Theorem 1. *[6] Let X be a random star-shaped set and $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of random star-shaped sets drawn from X. Then,*

(a) If $E(\|k_X\|) < \infty$ and $E(\|\rho_X\|_p) < \infty$, then
$\overline{k_X} \overset{a.s.-P}{\longrightarrow} E(k_X)$ and $\overline{\rho_X} \overset{a.s.-P}{\longrightarrow} E(\rho_X)$. Therefore, $\overline{X} \overset{a.s.-P}{\longrightarrow} E(X)$.

(b) If $E(\|k_X\|^2) < \infty$ and $E(\|\rho_X\|_p^2) < \infty$, then
$\widehat{\sigma}_{k_X}^2 \overset{a.s.-P}{\longrightarrow} \sigma_{k_X}^2$ and $\widehat{\sigma}_{\rho_X}^2 \overset{a.s.-P}{\longrightarrow} \sigma_{\rho_X}^2$. Therefore, $\widehat{\sigma}_X^2 \overset{a.s.-P}{\longrightarrow} \sigma_X^2$.

Example 1. Let X be a random rectangle-shaped set (a particular case of a random star-shaped set) so that the upper right vertex is generated by following real normal distributions of means 2 and 3, respectively, and variance equal to 1; the longest side is distributed as an $U(1,3)$ and the shortest one as an $U(3,5)$. A sample $\{X_i\}_{i=1}^{10}$ of rectangle-shaped sets i.i.d. as X is generated. The rectangles are centered on their center of gravity. The centered sample and the corresponding sample mean are represented in Figure 4.

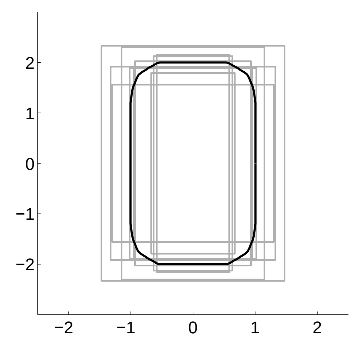

Fig. 4 Sample mean (in black) of a sample of 10 (gray-culoured) rectangles

It should be noticed that the sample mean is not a rectangle, as the corners are rounded due to the directional averaging.

The sample variance is computed for $\tau = 1$ (the sets are centered so that the importance is given to the imprecision) providing $\widehat{\sigma}_X^2 = \widehat{\sigma}_{\rho_X}^2 = .1192$.

4 Comparison of centered star-shaped sets

Let $A, B \in \mathcal{K}_s^*(\mathbb{R}^p)$ be two *centered* star-shaped sets (i.e., two star-shaped sets with common center which, without lack of generality, can be assumed to be 0) and let $\widetilde{\mathcal{K}_s^*}(\mathbb{R}^p)$ be the space of, either non-empty or empty, centered star-shaped sets of $\mathbb{R}^p$. In the same way that the so-called *length* of intervals and (fuzzy) sets has been used previously to develop statistics to compare convex and compact sets (see [21, 22]), the analogous concept can be considered for star-shaped sets. Thus, the *mean directional length* of A is defined in terms of the radial function by

$$S(A) = 2 \int_{\mathbb{S}^{p-1}} \rho_A(u)\ d\lambda_{\mathbb{S}^{p-1}}(u), \tag{7}$$

where $\lambda_{\mathbb{S}^{p-1}}$ denotes the normalized Lebesgue measure on the sphere. It should be noted that $S(A)$ is not the area of A, but an average of the magnitude of ρ_A over the unit sphere. It generalizes, in this way, the length of the intervals directionally, as it is always the case for the radial function.

Regarding the intersection, it is clear that $A \cap B \in \widetilde{\mathcal{K}_s^*}(\mathbb{R}^p)$. The *mean directional length of* $A \cap B$ can be expressed as follows:

$$S(A \cap B) = 2 \int_{\mathbb{S}^{p-1}} \min\left(\rho_A(u), \rho_B(u)\right)\ d\lambda_{\mathbb{S}^{p-1}}(u). \tag{8}$$

Based on the ideas in [23], the *degree of inclusion of* A *in* B, denoted by $Inc(A, B)$, is a value in $[0, 1]$ which can be defined by considering the quotient between the mean directional length of the intersection of A and B and the the mean directional length of the reference set A, i.e.

$$Inc(A, B) = \frac{S(A \cap B)}{S(A)}. \tag{9}$$

If A is included in B, then it is clear that $S(A \cap B) = S(A)$ and $Inc(A, B) = 1$; Otherwise, $S(A \cap B) < S(A)$ and $Inc(A, B) < 1$.

It is also possible to define the *degree of similarity of* A *and* B, denoted by $Sim(A, B)$, by following the ideas in [8], as the quotient between the shape of the intersection of A and B and the shape of the union of A and B, i.e.

$$Sim(A, B) = \frac{S(A \cap B)}{S(A \cup B)}, \tag{10}$$

where

$$S(A \cup B) = 2 \int_{\mathbb{S}^{p-1}} \max\left(\rho_A(u), \rho_B(u)\right)\ d\lambda_{\mathbb{S}^{p-1}}(u).$$

In this case, if A is equal to B, then $S(A \cap B) = S(A \cup B)$ and $Sim(A, B) = 1$; Otherwise, $S(A \cap B) < S(A \cup B)$ and $Sim(A, B) < 1$. Moreover, $Sim(A, B) < Inc(A, B)$ in all the situations.

Two illustrative examples concerning rectangle-shaped sets are shown in Figure 5. On the left part of the graphic, two partially overlapping centered rectangles A (in gray) and B (in black) are depicted. If we compute both the inclusion degree of A in B and the similarity degree between A and B, we obtain that $Inc(A, B) = .6822$ whereas $Sim(A, B) = .459$. On the right part of the graphic, the rectangle A (in gray) is completely contained in the rectangle B (in black). The computation of both indexes in this case leads us to the following results: $Inc(A, B) = 1$ and $Sim(A, B) = .5509$.

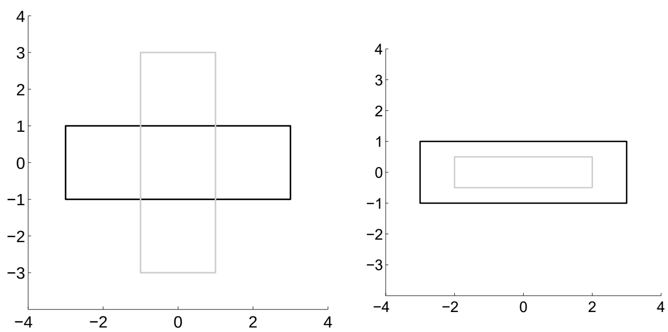

Fig. 5 Comparison of two rectangle-shaped sets in two different situations

The measures presented in this section might be greatly useful in the context of image processing. Therefore, it would be interesting to develop a deep statistical analysis about these measures in the near future.

5 Conclusions

An alternative representation for the class of star-shaped sets, called *center-radial characterization*, has been described. It has been shown to be useful for identifying intuitively the original shape of the sets. On the basis of this representation, some descriptive statistics for random star-shaped sets have been provided. Additionally, comparison measures based on the concept of mean directional length have been proposed. These measures are expected to be the starting point of an interesting research line in the area of image analysis. Furthermore, all the concepts provided in this work will be extended to the case of fuzzy subsets of $\mathbb{R}^p$ in a near future.

Acknowledgements

The research in this paper has been partially supported by MTM2013-44212-P, GRUPIN14-005 and the COST Action IC1408.

References

1. Artstein Z, Vitale R (1975) A strong law of large numbers for random compact sets. Ann Probab 3:879–882
2. Blanco-Fernández A, Corral N, González-Rodríguez G (2011) Estimation of a flexible simple linear model for interval data based on set arithmetic. Comp Stat Data An 55(9):2568–2578
3. Beresteanu A, Molchanov I, Molinari F (2011) Sharp identification regions in models with convex moment predictions. Econometrica 79:1785–1821
4. Choirat C, Seri R. (2013) Bootstrap Confidence Sets for the Aumann Mean of a Random Closed Set. Comp Stat Data Anal 71:803–817
5. Colubi A, González-Rodríguez G (2017) On Some Functional Characterizations of (Fuzzy) Set-Valued Random Elements. In: Ferraro MB, Giordani P; Vantaggi B, Gagolewski M, Gil MA, Grzegorzewski P, Hryniewicz O (eds) Soft Methods for Data Science. Series Advances in Intelligent Systems and Computing 456. Springer, Heidelberg
6. González-Rodríguez G, Ramos-Guajardo AB, Colubi A, Blanco-Fernández A (2017). A new framework for the statistical analysis of set-valued random elements. Submitted
7. Corral N, Gil MA, Gil P (2011) Interval and fuzzy-valued approaches to the statistical management of imprecise data. In: Pardo L, Balakrishnan N, Gil M (eds) Modern Mathematical Tools and Techniques in Capturing Complexity. Series Understanding Complex Systems. Springer, Heidelberg
8. Dubois D, Prade H (1980). Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York
9. Ferraro MB, Coppi R, González-Rodríguez G, Colubi A (2010) A linear regression model for imprecise response. Int J Approx Reason 51(7):759–770
10. Ghosh PK, Kumar KV (1998) Support function representation of convex bodies, its application in geometric computing, and some related representations. Comp Vis Image Unders 72(3):379–403
11. Gil MA, Montenegro M, González-Rodríguez G, Colubi A, Casals MR (2006) Bootstrap approach to the multi-sample test of means with imprecise data. Comp Statist Data Anal 51:148–162
12. Gil MA, González-Rodríguez G., Colubi A., Montenegro M. (2007) Testing linear independence in linear models with interval-valued data. Comp Statist Data Anal 51(6):3002–3015
13. González-Rodríguez G, Montenegro M, Colubi A, Gil MA (2006) Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst 157:2608-2613
14. González-Rodríguez G, Colubi A, Gil MA (2011) Fuzzy data treated as functional data: A one-way ANOVA test approach. Comp Statist Data Anal 56: 943–955
15. Körner R (2000) An asymptotic α-test for the expectation of random fuzzy variables. J Statist Plan Infer 83:331–346
16. Minkowski H (1896) Geometrie der Zahlen. Teubner, Leipzig
17. Miranda E, Couso I, Gil P (2005) Random sets as imprecise random variables. J Math Annal Appl 307(1):32–47

18. Molchanov I (2005) Theory of Random Sets. Springer, London
19. Puri ML, Ralescu DA (1985) Limit theorems for random compact sets in Banach space. Math Proc Cambridge Philos 97:151–158
20. Ramos-Guajardo AB, Lubiano MA (2012) K-sample tests for equality of variances of random fuzzy sets. Comp Statist Data Anal 56(4):956–966
21. Ramos-Guajardo AB, Colubi A, González-Rodríguez G (2014) Inclusion degree tests for the Aumann expectation of a random interval. Inform Sci 288:412–422
22. Ramos-Guajardo AB, Colubi A, González-Rodríguez G (2014) Inclusion and exclusion hypothesis tests for the fuzzy mean. Fuzzy Sets Syst 243:70–83
23. Sánchez E (1979) Inverses of fuzzy relations. Application to possibility distributions and medical diagnosis. Fuzzy Sets Syst 2:75–86
24. Schneider R (1993) Convex Bodies: the Brunn-Minkowski Theory. Cambridge University Press, Cambridge
25. Tonon F, Bernardini A (1998) A random set approach to the optimization of uncertain structures. Comp Struct 68:583–600

To Pedro, our research father and grandfather, a good man who encouraged us to follow this road. Thanks for giving us so much. Always in our hearts.