On Some Functional Characterizations of (Fuzzy) Set-Valued Random Elements

Ana Colubi and Gil Gonzalez-Rodriguez

Abstract One of the most common spaces to model imprecise data through (fuzzy) sets is that of convex and compact (fuzzy) subsets in $\mathbb{R}^p$. The properties of compactness and convexity allow the identification of such elements by means of the so-called support function, through an embedding into a functional space. This embedding satisfies certain valuable properties, however it is not always intuitive. Recently, an alternative functional representation has been considered for the analysis of imprecise data based on the star-shaped sets theory. The alternative representation admits an easier interpretation in terms of 'location' and 'imprecision', as a generalized idea of the concepts of mid-point and spread of an interval. A comparative study of both functional representations is made, with an emphasis on the structures required for a meaningful statistical analysis from the ontic perspective.

1 Introduction

The statistical analysis of (fuzzy) set-valued data from the so-called 'ontic' perspective has frequently been developed as a generalization of the statistics for interval data (see, e.g., [1]). From this 'ontic' perspective, (fuzzy) set-valued data are considered as whole entities, in contrast to the epistemic approach, which considers (fuzzy) set-valued data as imprecise measurements of precise data (see, e.g., [2]). Both the arithmetic and metric structure to handle this 'ontic' data is often based on an extension of the Minkowski arithmetic and the distance between either infima and suprema or mid-points and spreads for intervals. In this way, key concepts such as the expected value or the variability, are naturally defined as an extension of the classical notions within the context of (semi-)linear metric spaces.

The generalization of the concept of interval to $\mathbb{R}^p$ keeps the compactness and convexity properties, and this allows the identification of the contour of the convex

A. Colubi (✉) · G. Gonzalez-Rodriguez
Indurot and Department of Statistics, University of Oviedo, 33007 Oviedo, Spain
e-mail: colubi@uniovi.es

G. Gonzalez-Rodriguez
e-mail: gil@uniovi.es

M.B. Ferraro et al. (eds.), *Soft Methods for Data Science*, Advances
in Intelligent Systems and Computing 456, DOI 10.1007/978-3-319-42972-4_17

and compact sets in $\mathbb{R}^p$ by means of the support function (see, e.g., [6]). The support function is coherent with the Minkowski arithmetic, but sometimes this is not easy to interpret. In [4] the so-called kernel-radial characterization is investigated as an alternative to the support function based on a representation on polar coordinates. This polar representation is established in the context of the star-shaped sets, and is connected with the developments in [3]. It is coherent with alternative arithmetics and distances generalizing the concepts of location and imprecision in an intuitive way, which are of paramount importance in the considered context.

The aim is to show a comparative study of the support function and the kernel-radial representation through some examples. Methodological and practical similarities and differences of both representations for statistical purposes will be highlighted. The rest of the paper is organized as follows. In Sect. 2 both functional representations are formalized and their graphical visualization is shown for some examples. Section 3 is devoted to the comparison of the corresponding statistical frameworks. Section 4 finalizes with some conclusions.

2 The Support Function and the Kernel-Radial Characterization

Since the space of fuzzy sets to be considered is a level-wise extension of (convex and compact) sets, the analysis will focus on $\mathcal{K}_c(\mathbb{R}^p) = \{A \subset \mathbb{R}^p \mid A \neq \emptyset$, compact and convex$\}$. For any $A \in \mathcal{K}_c(\mathbb{R}^p)$, the support function of A is defined as $s_A : \mathbb{S}^{p-1} \to \mathbb{R}$ such that $s_A(u) = \sup_{a \in A} \langle a, u \rangle$ for all $u \in \mathbb{S}^{p-1}$, where $\mathbb{S}^{p-1}$ stands for the unit sphere in $\mathbb{R}^p$ and $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^p$. The support function s_A is continuous and square-integrable on $\mathbb{S}^{p-1}$ and characterizes the set A (see, e.g., [6]).

On the other hand, let $\mathcal{K}_S(\mathbb{R}^p)$ be the space of *star-shaped sets* of $\mathbb{R}^p$, i.e., the space of the nonempty compact subsets $A \subset \mathbb{R}^p$ so that there exists $c_A \in A$ such that for all $a \in A$, $\lambda c_A + (1-\lambda)a \in A$, for all $\lambda \in [0, 1]$, that is, all the points of A are 'visible' from c_A (see, e.g., [6]). The set of points $c_A \in A$ fulfilling the above condition is called *kernel* of A, ker(A). Each $c_A \in ker(A)$ is considered a *center* of A. Obviously, $\mathcal{K}_S(\mathbb{R}) = \mathcal{K}_c(\mathbb{R})$, but for $p > 1$, $\mathcal{K}_c(\mathbb{R}^p) \subset \mathcal{K}_S(\mathbb{R}^p)$.

A star-shaped set A can be characterized by a center k_A (e.g., the center of gravity of the kernel), and the radial function defined on the unit sphere. The radial function identifies the contour by means of the distance to that center, i.e., by means of the polar coordinates (see, e.g., [6]). Formally, the center of gravity is given by the expected value of the uniform distribution on ker(A), that is, $k_A = \int_{\ker(A)} x d\mu_k$, being μ_k the normalized Lebesgue measure on ker(A). The *radial function* is defined as the mapping $\rho_A : \mathbb{S}^{p-1} \to \mathbb{R}^+$ such that $\rho_A(u) = \sup\{\lambda \geq 0 : k_A + \lambda u \in A\}$.

The radial function is the inverse of the gauge function, which has been used in [3] in the context of fuzzy star-shaped sets. However, in [3] the gauge function was not used as a basis for the arithmetic and the metric structure of the space, but in

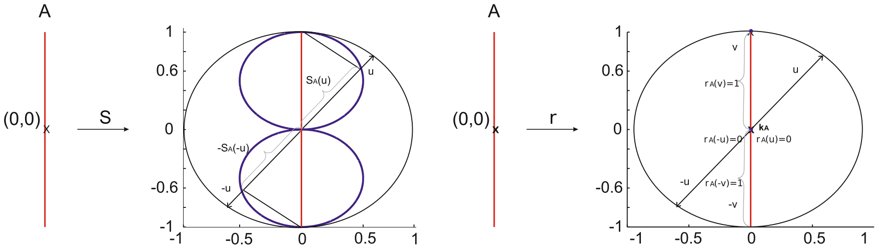

Fig. 1 Graphical representation of the support function (*left*) and the radial function (*right*) of a line in $\mathcal{K}_c(\mathbb{R}^2)$

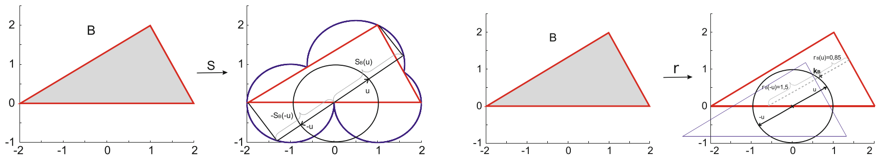

Fig. 2 Graphical representation of the support function (*left*) and the radial function (*right*) of a *triangle* in $\mathcal{K}_c(\mathbb{R}^2)$

combination with the usual structures, which has reduced the practical usefulness of the proposal.

In order to compare the interpretation of the support and the radial function, Figs. 1 and 2 show a graphical representation of both functions corresponding to a line and a triangle respectively. Since the characterizing functions are defined over the unit sphere, the representations show how each element of the unit sphere relates to the corresponding value. For the support function the sets in $\mathbb{R}^2$ are projected on each one of the directions of the unit sphere and the maximum is computed. In this way, the support function is the distance from the center to the contour of the blue lines. Although this identifies in a unique way the boundaries of the set, the result is not easy to relate with the original shape at first glance. The radial function represents the polar coordinates of the contour line of the original set, that is the radius to each point from the pole (i.e. the steiner point of the kernel). Consequently, the shape of the radial function is straightforwardly connected with the original shape.

For the radial representation, k_A is a center of A, describing the location of the set, and ρ_A shows how far the contour line is from this center pointwise. Thus, in line with the idea of mid-point (location) and spread (imprecision) of an interval, k_A and the radial function ρ_A can be identified with the generalized location and imprecision of a star-shaped set respectively.

A previous attempt was made to define generalized concepts of location and imprecision on the basis of the support function by considering the so-called mid-spread representation [7]. This representation is so 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 $u \in \mathbb{S}^{p-1}$. That is, the generalized mid-point/spread is connected with the location/imprecision associated with each direction. This fact entails an interpretational profit, but also some drawbacks from an operational view. Moreover, it inherits the visualization shortcomings from the support function. The main problem is that it is disfficult to determine when a function $s : \mathbb{S}^{p-1} \to \mathbb{R}$ is a support function of any $A \in \mathcal{K}_c(\mathbb{R}^p)$, and this is translated to the mid-function. This problem, however, does not affect the kernel-radial representation, because any function $\rho_A : \mathbb{S}^{p-1} \to \mathbb{R}^+$ is a radial function of a given set.

3 Statistical Frameworks

Either through the support function or through the kernel-radial characterization, the space of the corresponding set-valued elements can be embedded into a Hilbert space, namely, $\mathcal{H}_s = L^2(\mathbb{S}^{p-1})$ endowed with the normalized Lebesgue measure on $\mathbb{S}^{p-1}$, λ_p, for the case of the support function and $\mathcal{H}_r = \mathbb{R}^p \times L^2(\mathbb{S}^{p-1})$ endowed with $\mu_p \times \lambda_p$ for the case of the kernel-radial characterization. Nevertheless, in order to have a meaningful embedding useful for statistical purposes, the arithmetic and metric structures of the original spaces and the Hilbert ones should agree.

It is well known that the support function transfers the Minkowski arithmetic into $\mathcal{H}_s$ and, with the proper metrics, it makes $\mathcal{K}_c(\mathbb{R}^p)$ isometric to a cone of $\mathcal{H}_s$. This arithmetic is defined so that $A +_M \tau B = \{a + \tau b \,|\, a \in A, b \in B\}$ for all $A, B \in \mathcal{K}_c(\mathbb{R}^p)$ and $\tau \in \mathbb{R}$, and verifies that $s_{A+_M \tau B} = s_A + \tau s_B$ for all $\tau \geq 0$. The Minkowski addition is not always meaningful, and there exist various alternatives (see, e.g., [5]).

When the sets are characterized in terms of kernel-radial elements, the natural arithmetic should be coherent as well, that is, $A +_r \tau B$ should be the element in $\mathcal{K}^*(\mathbb{R}^p)$ such that $k^k_{A+_r \tau B} = k_A + \tau k_B$ and $\rho_{A+_r \tau B} = \rho_A + \tau \rho_B$, where the $+$ operator denotes either the usual sum of two points in $\mathbb{R}^p$ or the usual sum of two functions in $L^2(\mathbb{S}^{p-1})$, respectively, for all $A, B \in \mathcal{K}(\mathbb{R}^p)$ and $\tau \in \mathbb{R}$.

Figure 3 shows how sometimes the kernel-radial arithmetic may be more useful than Minkowski's one. The Minkowski and the kernel-radial sum of two lines is shown graphically. The Minkowski sum of two elements in $\mathcal{K}_c(\mathbb{R}^2)$ with null area in $\mathbb{R}^2$ and the same shape results in a convex set with different shape and non-null area. On the contrary, the kernel-radial arithmetic keeps the shape and the surface of the sets.

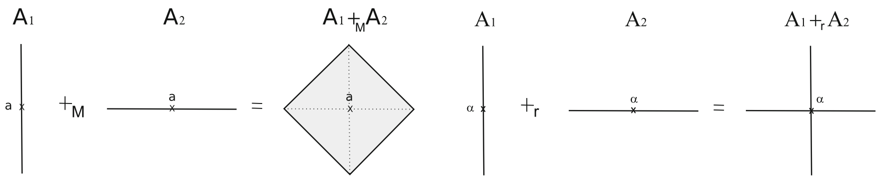

Fig. 3 Minkowski (*left*) and radial (*right*) sum of two segments

Concerning the metric structure, L^2-type metrics are normally considered for statistical purposes. For instance, for the support function-related characterizations, it is common to consider the generalized family for $\theta \in [0, +\infty)$

$$d_\theta(A, B) = \sqrt{||\text{mid}_A - \text{mid}_B||_p^2 + \theta||\text{spr}_A - \text{spr}_B||_p^2},$$

for all $A, B \in \mathcal{K}_c(\mathbb{R}^p)$ where $||\cdot||_p$ is the usual L^2-type norm for functions defined on $\mathbb{S}^{p-1}$ with respect to λ_p [7]. In an analogous way, for the kernel-radial representation, the natural family of metrics for statistical purposes from an ontic point of view is

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

for all $A, B \in \mathcal{K}(\mathbb{R}^p)$ and $\tau \in (0, 1)$, where $||\cdot||$ is the usual Euclidean norm in $\mathbb{R}^p$.

With these structures, it is clear that the considered spaces can be identified with cones of Hilbert spaces, and all the statistical concepts and tools defined in general Hilbert space apply in this context, taking into account that some constraints may arise whenever it is required to remain in the cone. Thus, notions such as random element, expected value, variance or covariance operator, and basic results, such as the CLT, are directly inherited from the theory in Hilbert spaces in the same way for both characterizations. The unique methodological difference in this respect is that, although it is trivial to check if a radial function remains in the cone (i.e. $\rho_A(u) \geq 0$ for all $u \in \mathbb{R}^p$), this is not the case for the support function.

4 Conclusions

The support function has traditionally been used to characterize compact and convex sets. This is specially useful when the Minkowski arithmetic is suitable. We have shown that this concept is not always intuitive. As an alternative, the kernel-radial representation is proposed. One of the main advantages of this new representation is that it is easy to interpret in terms of generalized concepts of mid-spread for intervals. The statistical analysis involving both kind of elements can be reduced in both cases to the Hilbert case, so no specific methodology is required to be developed for many common problems. Moreover, the characterization of the cone where the sets are embedded is trivial and similar to the interval case (i.e., non-negativity constraints). This entails a substantial methodological simplification when it is essential to guarantee that any element remains in the cone. Concerning the arithmetic, it has been shown that the Minkowski sum is not always suitable when $p > 1$, as it does not keep shapes or areas, while the arithmetic based on the kernel-radial representation can be a suitable alternative for cases where that is important.

All the discussions in this paper can be extended to the case of fuzzy sets by considering levelwise-defined concepts. Namely, let $\mathcal{F}_c(\mathbb{R}^p)$ be the space of fuzzy sets $U : \mathbb{R}^p \to [0, 1]$ whose α-level sets $U_\alpha \in \mathcal{K}_c(\mathbb{R}^p)$ for all $\alpha \in (0, 1]$. Then, the support function can be defined as $s_A : \mathbb{S}^{p-1} \times (0, 1] \to \mathbb{R}$ so that $s_A(u, \alpha) = s_{A_\alpha(u)} \sup_{a \in A_\alpha} \langle a, u \rangle$ for all $u \in \mathbb{S}^{p-1}$, and $\alpha \in (0, 1]$. In the same way, the Minkowski arithmetic is level-wise defined, and the metric is established wrt the joint normalized Lebesgue measure on $\mathbb{S}^{p-1} \times (0, 1]$. Analogous developments can be performed for the case of the kernel-radial representation. The unique technical burden that distinguishes the case of fuzzy sets from the case of standard sets is the problem of building a fuzzy set from the functions on the respective Hilbert spaces, if possible, but this can be done by taking into account the well-known properties that guarantee that a set of indexed levels $\{A_\alpha\}_{\alpha \in [0,1]}$ determines a fuzzy set.

Acknowledgments The research in this paper has been partially supported by MTM2013–44212-P, GRUPIN14-005 and the COST Action IC1408. The authors would also like to thank Dr. Angela Blanco for her scientific support.

References

1. Colubi A, Gonzalez-Rodriguez G (2015) Fuzziness in data analysis: towards accuracy and robustness. Fuzzy Sets Syst 281:260–271
2. Couso I, Dubois D (2014) Statistical reasoning with set-valued information: ontic vs. epistemic views. Int J Approximate Reasoning 55:15021518
3. Diamons P (1990) A note on fuzzy starshaped fuzzy sets. Fuzzy Sets Syst 37:193–199
4. Gonzalez-Rodriguez G, Blanco-Fernandez A, Colubi A (2015) A new framework for the statistical analysis of set- and fuzzy set-valued random elements. Submitted
5. Molchanov I (1998) Averaging of random sets and binary images. Quarterly 11:371–384
6. Schneider R (1993) Convex bodies: the Brunn-Minkowski theory. Cambridge University Press, Cambridge
7. Trutschnig W, Gonzalez-Rodriguez 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