Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A new framework for the statistical analysis of set-valued random elements ☆

Gil Gonzalez-Rodriguez, Ana Belen Ramos Guajardo, Ana Colubi *, Angela Blanco

INDUROT/Dept. of Statistics and OR, Campus de Mieres, University of Oviedo, 3600 Mieres, Spain

ARTICLE INFO

Article history:
Received 10 January 2017
Received in revised form 17 October 2017
Accepted 18 October 2017
Available online xxxx

Keywords:
Imprecise data
Set-valued data
Support function
Star-shaped set
Functional characterizations
Center-radial representation
Location/imprecision
Arithmetic
Metrics
Separable Hilbert space
Statistics
Random set

ABSTRACT

The space of nonempty convex and compact (fuzzy) subsets of $\mathbb{R}^p$, $\mathcal{K}_c(\mathbb{R}^p)$, has been traditionally used to handle imprecise data. Its elements can be characterized via the support function, which agrees with the usual Minkowski addition, and naturally embeds $\mathcal{K}_c(\mathbb{R}^p)$ into a cone of a separable Hilbert space. The support function embedding holds interesting properties, but it lacks of an intuitive interpretation for imprecise data. As a consequence, it is not easy to identify the elements of the image space that correspond to sets in $\mathcal{K}_c(\mathbb{R}^p)$. Moreover, although the Minkowski addition is very natural when $p = 1$, if $p > 1$ the shapes which are obtained when two sets are aggregated are apparently unrelated to the original sets, because it tends to convexify. An alternative and more intuitive functional representation will be introduced in order to circumvent these difficulties. The imprecise data will be modeled by using star-shaped sets on $\mathbb{R}^p$. These sets will be characterized through a center and the corresponding polar coordinates, which have a clear interpretation in terms of *location* and *imprecision*, and lead to a natural directionally extension of the Minkowski addition. The structures required for a meaningful statistical analysis from the so-called ontic perspective are introduced, and how to determine the representation in practice is discussed.

1. Introduction

For the last decades the statistical analysis of imprecise-valued random variables has awakened a great interest from both the epistemic and the ontic viewpoints (see, for extensive comparative discussions, e.g., [4,7]). These random variables associate outcomes of a random experiment (modelled through probability spaces) with elements in generalized spaces, such as the space of compact real intervals, the space of convex, and compact subsets of $\mathbb{R}^p$, the space of fuzzy numbers, or the space of convex and compact p-dimensional fuzzy sets. From the 'ontic' perspective, the one considered in this paper, (fuzzy) set-valued data are regarded as whole entities (see, e.g., [2,4–6]), in contrast to the epistemic approach, which deals with (fuzzy) set-valued data as imprecise measurements of precise data (see, e.g., [7,8,11,17]).

The elements of the above-mentioned spaces are generally parametrized by vectors/functions to derive more operational statistical techniques. For instance, any real compact interval can be characterized by its contour or endpoints

☆ This paper is part of the Virtual special issue on Soft methods in probability and statistics, Edited by Barbara Vantaggi, Maria Brigida Ferraro, Paolo Giordani.

* Corresponding author.
E-mail address: colubi@uniovi.es (A. Colubi).

https://doi.org/10.1016/j.ijar.2017.10.025
0888-613X/© 2017 Published by Elsevier Inc.

(infimum/supremum), or by its midpoint (related to the location) and spread (related to the imprecision). For the space of fuzzy numbers, each level set is a real compact interval, so the corresponding level-wise characterization is usually considered. Zadeh's extension principle, which agrees with the Minkowski arithmetic, is fully meaningful and has been extensively used for statistical purposes (see, e.g., [9,16,19]). For the p-dimensional case, the contour of the compact and convex (fuzzy) sets is usually identified by the support function (see, e.g., [10,22]). By operating with it, the location/imprecision can also be characterized with the so-called generalized mid-spread representation [26] or alternative characterizations based on the Steiner point and shape deviations expressed in terms of the support function (see, e.g., [1,23] and references therein). Nevertheless, neither the support function, nor the elements of the related representations, have a fully intuitive meaning, even if theoretically they provide a valid characterization. Moreover, the convexification property of the Minkowski arithmetic also makes it difficult to relate the meaning of two aggregated compact and convex sets with the original shapes when $p > 1$. These shortcomings are discussed for the set-valued case and illustrated through examples. An alternative functional representation based on the theory of star-shaped sets (see, e.g., [22]) is proposed, and the foundations to develop statistics in this new framework are set up.

The rest of the manuscript is structured as follows. In Section 2 the current paradigm based on the support function is formally introduced and some examples showing its lack of interpretability in some cases are discussed. In Section 3 a new parametrization is considered, based on a point related to the location, and a polar function, related to the imprecision. The main properties of this characterization are discussed. The new framework considers star-shaped sets as a natural setting, but it includes the usual convex and compact sets as particular cases. Examples to illustrate the new parametrization are provided. Section 4 introduces a pre-processing step to general star-shaped sets in order to establish their location in practice in a robust way. Finally, Section 5 summarizes some conclusions and future research.

2. Current framework

Let $\mathcal{K}_c(\mathbb{R}^p)$ be the space of nonempty compact and convex subsets of $\mathbb{R}^p$. This space is normally endowed with the Minkowski arithmetic, which generalizes the standard interval arithmetic as follows:

$$A + \gamma B = \{a + \gamma b \,|\, a \in A, b \in B\} \text{ for all } A, B \in \mathcal{K}_c(\mathbb{R}^p) \text{ and } \gamma \in \mathbb{R}.$$

Given $A \in \mathcal{K}_c(\mathbb{R}^p)$, the well-known support function characterizes the contour of A (see, e.g., [22]). It is defined as $s_A : \mathbb{S}^{p-1} \to \mathbb{R}$ such that

$$s_A(u) = \sup_{a \in A} \langle a, u \rangle \text{ for all } u \in \mathbb{S}^{p-1},$$

where $\mathbb{S}^{p-1}$ denotes the unit sphere in $\mathbb{R}^p$ and $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^p$ with associated norm $\| \cdot \|$. The support function s_A is continuous and square-integrable on $\mathbb{S}^{p-1}$.

From now on, given $a \in \mathbb{R}^p$ and $\epsilon > 0$, $B(a, \epsilon)$ will denote the open ball centered at a and with radius ϵ, that is

$$B(a, \epsilon) = \{x \in \mathbb{R}^p \,|\, \|a - x\| < \epsilon\},$$

and $\overline{B}(a, \epsilon)$ will denote the corresponding closed ball.

The space $\mathcal{K}_c(\mathbb{R}^p)$ can be embedded into the separable Hilbert space of the square integrable functions $\mathcal{H} = \mathcal{L}^2(\mathbb{S}^{p-1}, \zeta)$, with ζ the normalized Lebesgue surface measure on $\mathbb{S}^{p-1}$, by means of the mapping $s : \mathcal{K}_c(\mathbb{R}^p) \to \mathcal{L}^2(\mathbb{S}^{p-1})$ defined by $s(A) = s_A$. The support function is not linear, but semi-linear, i.e.

$$s_{A+\gamma B}(u) = s_A(u) + |\gamma| s_B(\text{sign}(\gamma) u),$$

for all $A, B \in \mathcal{K}_c(\mathbb{R}^p)$, $\gamma \in \mathbb{R}$ and $u \in \mathbb{S}^{p-1}$. Thus, s preserves the Minkowski addition and product by non-negative scalars, and it makes $\mathcal{K}_c(\mathbb{R}^p)$ to be isomorph to a cone of $\mathcal{L}^2(\mathbb{S}^{p-1}, \zeta)$ by using the induced L_2 distance on $\mathcal{K}_c(\mathbb{R}^p)$. The embedding s allows the development of statistical analyses in $\mathcal{K}_c(\mathbb{R}^p)$ by applying powerful techniques for Hilbert spaces (see, for instance, [15]). It should be underlined that once the support functions are used to characterize compact convex sets, the different developments refer to equivalence classes, since the elements in $\mathcal{L}^2(\mathbb{S}^{p-1}, \zeta)$ represent classes of functions that are equal almost surely.

Unfortunately, the support function of a convex set is not easy to visualize. In general, a graphical representation of s_A seems to be unrelated with the shape of A, as it is illustrated in the following simple example.

Example 1. Fig. 1 illustrates the computation of the support function of the rectangle drawn in green color, centered at (0,0) and with corners (±10,±1). The unit sphere is the solid black line circumference. For any fixed point u in the unit sphere, the rectangle is orthogonally projected on the line passing by the origin and with direction given by u (dotted black line), leading to an interval over this line. The value of the support function at u is the maximum value of this projection in the direction given by u (the length of the red segment in this particular case). Given any u in the unit sphere, the function drawn in blue represents the point where such a maximum is attained, and the distance to the origin corresponds to the value of the support function. This representation can only be done if the origin belongs to the set, otherwise there will be directions in the unit sphere for which the support function becomes negative. The graphic is complemented with the representation of the support function in the sphere parametrized by angles in $[0, 2\pi)$ (Fig. 2).

Please cite this article in press as: G. Gonzalez-Rodriguez et al., A new framework for the statistical analysis of set-valued random elements, Int. J. Approx. Reason. (2017), https://doi.org/10.1016/j.ijar.2017.10.025