A new bridge between fuzzy sets and statistics

Pedro Terán

Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
E-33071 Gijón (Spain)
Email: teranpedro@uniovi.es

Abstract— *Our basic question is as follows: given a, maybe multivariate, probability distribution, how to formalize the idea that some points are more central than others? We present an account of the notion of centrality which is based on fuzzy events and is valid for single distributions and for families of distributions (including imprecise probability models).*
This unifying framework is natural and subsumes many known concepts like the following (not an exhaustive list): (a) univariate location estimators like the mean, the median and the mode, (b) the interquartile interval and the Lorenz curve of a random variable, (c) several generalized medians, trimmed regions and statistical depth functions from multivariate analysis, (d) most known location estimators for random sets, (e) the probability mass function of a discrete random variable and the coverage function of a random closed set, (f) the Choquet integral with respect to an infinitely alternating or infinitely monotone capacity.

Keywords— Centrality, Fuzzy event, Multivariate analysis, Random set, Statistical depth.

1 Introduction

All points are central, but some points are more central than others.[1]

Our aim is to show how a simple idea with a natural interpretation within fuzzy set theory allows to define, in a unified way, many notions from statistics. These include well-known location estimators for random variables; some multivariate generalizations which grew into the theory of statistical depth functions in the last fifteen years; trimming methods based on them; means and medians of random sets; and even Choquet integrals. In fact, the notion applies as soon as we have a model to which we are able to associate a family of probability distributions.

The main tool is the notion of a fuzzy event and its probability [25]. A *fuzzy event* in a measurable space Ω is a measurable mapping $A : \Omega \to [0, 1]$. The membership $A(\omega)$ denotes the degree to which event A occurs when $\omega \in \Omega$ occurs. A probability measure on Ω extends to all fuzzy events by the natural formula

$$P(A) = \int A \mathrm{d}P.$$

The structure of this communication is as follows. Section 2 presents the fuzzy notion of centrality of a point in a probability distribution. Section 3 extends it to families of distributions. Sections 4, 5, 6 and 7 exhibit specific examples covered by our framework. Finally, Section 8 closes the paper with some final remarks.

2 Centrality and location estimators based on fuzzy events

What does it mean to speak about the 'centrality' of a point in a data set or a probability distribution? Although it is informally clear that the mean, the median and the mode are three traditional answers to the problem of pinpointing where the 'center' of a distribution is, a formal definition of centrality is far more elusive. In this section, an intuitively appealing view of centrality based on fuzzy set theory is presented.

We begin by choosing a family $\mathcal{A}$ of fuzzy events, called *reference events*. Roughly speaking, a point x is considered *central* if the probability of a reference event is high whenever it contains x. That is, *x is central if the probable reference events are implied by the occurrence of x*, whereas x is not central if it belongs to unlikely reference events. Note that this definition of centrality is relative to the chosen family of reference events.

This notion is reminiscent of the mode as the most likely point, and actually the mode is what appears when the reference events are chosen to be the (crisp) points.

It is possible to derive a crisp definition of centrality from the ideas above: we could say that a point is central if

$$P(A) \geq A(x) \quad \forall A \in \mathcal{A}.$$

But although this might capture the notion that the value of $P(A)$ is large insofar as the degree to which x belongs to A is large, a gradual restriction seems even more faithful to the idea. Thus, we say that a point is *central to degree at least* α (with respect to the family $\mathcal{A}$) if

$$P(A) \geq \alpha \cdot A(x) \quad \forall A \in \mathcal{A}.$$

The set of all the central points to degree at least α is the α-cut of a fuzzy set. That allows us to define the fuzzy set C given by

$$C(x) = \sup\{\alpha \in (0, 1] \mid \forall A \in \mathcal{A}, P(A) \geq \alpha A(x)\},$$

whose membership function indicates the degree of centrality of each point in the probability distribution P.

This definition can be rewritten in the language of fuzzy logic, as follows.

[1] Paraphrasing a famous sentence from Orwell's 'Animal farm'.

Proposition 1. *Denote Goguen's fuzzy implication by I. Then,*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), P(A)).$$

Since $P(A)$ can be seen as the degree of truth of the proposition 'A is probable' in an adequate fuzzy logic (see e.g. [7]), the fuzzy set C captures the intuitive idea that its elements are those such that *'for any reference event A, if $x \in A$ then A is probable'*.

In principle, the Goguen implication can be replaced by a different implication. In this paper, we content ourselves with showing that many statistical objects arise from this choice (whose reason is historical, as explained in the final section), without analyzing other possible choices. However, note that changing the implication may lead to obtaining the same examples with a mere modification of the shape of the reference events.

From a statistical point of view, those points with the maximal centrality can serve as location estimators which might be called *maximal centrality estimators*. In general, there may be more than one maximally central point (like e.g. several modes), so all the maximally central points form the *central region*. Points with centrality at least α form the *α-central region*.

3 Extension to families of distributions

As shown in Section 4, the notion of centrality presented above is directly related to some traditional location estimators, as well as to several multivariate generalizations which have been introduced and studied mostly in the last fifteen years. However, to account for some of those cases and specially to cover more general types of models like random sets, it is necessary to extend it to study the centrality of a point in a *family* of distributions.

When the model consists of more than one single distribution, the underlying distribution generating the data is known or assumed to belong to a specified family. Some situations covered by this more general setting are the following.

(a) A parametric or non-parametric model.

(b) Upper and lower probabilities.

(c) A Choquet capacity.

(d) A random set.

(e) A credal set.

(f) A neighbourhood of a probability.

(g) Other families defined in terms of the distribution of a random variable.

Given a random variable ξ, instances of (g) include the family $\{P_{g \cdot \xi} \mid g \text{ is } [0,1]\text{-valued}\}$ used to define the zonoid [11] of ξ; and the family $\{Q \leq \alpha^{-1} P_\xi\}$ used to define trimmed regions in [4]). In such examples, a family of distributions is constructed whose central region contains interesting information about ξ, although maybe no longer related to centrality itself. For instance, the lift zonoid of ξ, a multivariate generalization of the Lorenz curve, is the central region of a family of distributions defined from the distribution of ξ.

In such a that situation, for statistical purposes a point should be considered central in the family if it is potentially a central point of the true distribution, which could be any member of the family. The rationale is that a model defined by a family of distributions is more imprecise than pinpointing a single distribution, so the compatibility of a point being central with the information available should be larger [2].

That suggests the following definition: given a family $\mathcal{P}$ of probability distributions, the fuzzy set C of central points is defined by

$$C(x) = \sup\{\alpha \in (0,1] \mid \forall A \in \mathcal{A}, \sup_{P \in \mathcal{P}} P(A) \geq \alpha A(x)\}.$$

The rest of the definitions (α-central region, etc.) are modified analogously.

Observe that the location information in a family of distributions is summarized using the same object as for a single distribution. Thus, the complexity of that summary is the same as for a single distribution (although calculating C may be computationally more expensive). That implies that statistical procedures based on C, devised for single distributions, will extend immediately to the situations above, particularly imprecise probability models (Choquet capacities, random sets).

As will be shown later, in the cases when C coincides with a statistical depth function such procedures already exist.

Let us finally note that some of the results below may require, for technical reasons, that the family $\mathcal{P}$ considered is not an arbitrary one but compact with respect to the topology of weak convergence (convergence in distribution).

4 Univariate examples

As already mentioned, the three basic examples of location estimators arise as special cases of maximally central points of a distribution.

Theorem 2. *Let ξ be a random variable. For appropriate choices of $\mathcal{A}$ and $\alpha \in (0,1]$, the α-central region C_α is:*

(1) The expectation of ξ, if it exists (otherwise, the corresponding α-central region is empty).

(2) The median of ξ.

(3) The mode of ξ, if ξ is discrete.

The necessary choices of α in Theorem 2 are: $\alpha = 1$ in (1); $\alpha = .5$ in (2); α the height of C in (3). In fact, in all the three cases α coincides with the height of the fuzzy set C. In subsequent examples, for ease of presentation we omit the choice of α in each case.

[2] In any case, it is also interesting to consider the dual approach: to which degree a point is guaranteedly central in *all* the distributions of the model. That was pointed out to me by Didier Dubois.

As to the other ingredient, $\mathcal{A}$, one may ask whether the families of reference events needed to retrieve these classical estimators may be very complicated. In fact, they are quite simple:

(1) $\{A_y(x) = |x-y|/(1+|x-y|)\}_{y \in \mathbf{R}}$ for the expectation.

(2) $\{A_y^+(x) = I_{[y,\infty)}(x), A_y^-(x) = I_{(-\infty,y]}(x)\}_{y \in \mathbf{R}}$ for the median.

(3) $\{A_y(x) = I_{\{y\}}(x)\}_{y \in \mathbf{R}}$ for the mode.

In (2) and (3), I_B denotes the indicator function of a crisp set B.

Subsequent examples use the same choices of $\mathcal{A}$ or suitable multivariate generalizations. Again to avoid burdensome presentation, we omit the specific choice of $\mathcal{A}$ in each case.

Interestingly, (1) is associated to a new characterization of the expectation of a random variable in terms of the bounded metric $d(x,y) = |x-y|/(1+|x-y|)$.

Proposition 3. *Let ξ be an integrable random variable. Then, its expectation is the unique real number x such that, for any number y,*

$$d(x,y) \leq Ed(\xi, y);$$

namely $E\xi$ is identified by the property that every number is closer in d-distance to $E\xi$ than it is (in mean) to ξ.

Proof (outline of the main ideas). There exists a characterization of the mean, due to Shafik Doss [6], as the unique real number x such that, for any y,

$$\|x - y\| \leq E\|\xi - y\|.$$

It suffices to show that both properties are equivalent. The easy part, with the help of the Jensen inequality, is to show that a number satisfying the condition in the statement must also fulfil Doss's property (and so our condition is formally stronger and only the mean could satisfy it).

To prove that the mean has indeed the property in the statement we need to go against Jensen's inequality, so things get more involved. The key is to take a divergent sequence of values for y and show that, when y is very far from x, Jensen's inequality is almost an equality, so in the limit we can 'reverse' it. □

Actually, this geometric characterization is valid in any separable Banach space, if we just replace the Euclidean distance by the norm.

A couple more univariate examples are as follows. These are no longer related to location estimation.

Proposition 4. *Let ξ be a random variable. For appropriate choices of $\mathcal{A}$ and $\alpha \in (0,1]$, the α-central region C_α is:*

(1) The interquartile interval of ξ.

(2) The Lorenz curve of ξ (more exactly, the curve together with the region below it, its hypograph).

Both the interquartile interval and the median are examples of the interval between the p-th quantile and the $(1-p)$-th quantile (for $p = .25$ and $p = .5$). Modulating the value of α (precisely, $\alpha = .5 - p$), we can obtain any such interval for $0 \leq p \leq .5$. For $p = 0$, that is the interval between the minimum and maximum values of ξ.

In this connection, let us remark that two known tools in exploratory data analysis, the box-plot and its multivariate generalization (the bag-plot), are plotted using only information subsumed by our framework.

5 Multivariate examples

In the multivariate case, the absence of a natural order makes it impossible for a point to have all the usual properties of the univariate median. That has led to the definition of many *depth functions* [14]; often the point maximizing the depth function is called a *generalized median* or a *maximal depth estimator*.

Depth functions aim at ordering, in a center-outward sense, the points in a data set or, more generally, those of $\mathbf{R}^d$ with respect to a given distribution. A typical application of depth (among many others) is outlier detection: sample points with very small depth are trimmed out as outliers.

The notion of depth is still somewhat vague, and its semantics is specially unclear for multimodal or very asymmetric distributions. Zuo and Serfling [26] underlined four desirable properties of depth functions, which had already been studied in particular cases:

(a) Depth is maximal at a center of symmetry of the distribution, if the latter exists.

(b) Depth decreases along any ray departing from the center.

(c) Depth vanishes as the distance to the center goes to infinity.

(d) The depth function should be affinely equivariant (so that conclusions do not depend on the chosen coordinate system).

It must be emphasized that, often, accepted notions of statistical depth fail some of those properties (e.g. L^p-depth).

From any depth function, an *α-trimmed region* can be constructed which is formed by those points having depth at least α. Plotting the contours of the α-trimmed regions constitutes an easy device for multivariate exploratory analysis.

To give the reader a grasp of how depth functions work, let us just mention some ways to calculate depth.

Convex hull peeling depth of a point in a data sample is calculated as follows. Take the convex hull of the sample. The points outside have depth 0. Now the sample points in the boundary of the convex hull are 'peeled off'; points left outside the cloud by doing so have depth proportional to 1. The sample points in the boundary of the convex hull of the remaining points are peeled off, and points left outside the cloud are given depth proportional to 2. The procedure goes on until the innermost points have been peeled off.

Simplicial depth of a point is calculated as the probability that it belongs to the random simplex whose vertices are independent and identically distributed copies of the variable.

Similarly, expected convex hull depth is inversely proportional to the number k of copies of the variable which are needed to ensure that the point is in the Aumann expectation of the sample $\{\xi_1, \cdots, \xi_k\}$.

Finally, for halfspace depth one considers the probability that the variable lies in a halfspace, and calculates the infimum of such values over those halfspaces whose boundary contains the point.

Some statistical depth functions in the literature are fuzzy sets of central points, as the following result shows.

Theorem 5. *Let ξ be a random vector. For appropriate choices of $\mathcal{A}$, $\mathcal{P}$ and $\alpha \in (0,1]$, the centrality $C(x)$ is:*

(1) The convex hull peeling depth of x (Huber [9], 1972; Barnett [2], 1976).

(2) The simplicial depth of x (Liu [15], 1990).

(3) The majority depth of x (Liu–Singh [16], 1993).

(4) The probability of the event $\{\xi = x\}$, if ξ is discrete.

The univariate examples presented above extend to the multivariate setting as follows (including the central regions associated to further notions of depth).

Theorem 6. *Let ξ be a multivariate random vector. For appropriate choices of $\mathcal{A}$, $\mathcal{P}$ and $\alpha \in (0,1]$, the α-central region C_α is:*

(1) The expectation of ξ, if it exists.

(2) The generalized median of ξ according to any of the depth functions (1-3) above.

(3) The α-trimmed region of ξ according to any of the depth functions (1-3) above.

(4) The generalized median and the α-trimmed region of ξ according to the halfspace depth (Tukey [23], 1975).

(5) The generalized median and the α-trimmed region of ξ according to the zonoid depth (Mosler [17], 2002).

(6) The generalized median and the α-trimmed region of ξ according to the symmetric order depth (Cascos–López-Díaz [5], 2004).

(7) The generalized median and the α-trimmed region of ξ according to the expected convex hull depth (Cascos [3], 2007).

(8) The mode of ξ, if ξ is discrete.

(9) The zonoid of ξ (see e.g. [11, 17]).

(10) The lift zonoid of ξ (Koshevoy–Mosler [12], 1998).

The lift zonoid generalizes the Lorenz curve and so serves to study multivariate concentration and inequality. It is also remarkable that it characterizes the distribution of ξ.

6 Multivalued examples

When a random set is considered, the statistical model takes the following form. The underlying random variable ξ is not directly observed; only a larger set of values X containing ξ is observed. Then, we know that ξ is such that $\xi \in X$ almost surely, namely ξ is a *selection* of X. In that case, $\mathcal{P}$ is the family of all selections of X.

Most location estimators for random sets are unified as special cases of central or α-central regions.

Theorem 7. *Let X be a random compact set and $\mathcal{P}$ the family of all its selections. Then, for appropriate choices of $\mathcal{A}$ and $\alpha \in (0,1]$, the α-central region C_α is:*

(1) The Aumann expectation of X (Kudō [13], 1954; Aumann [1], 1965) if Ω is non-atomic or X is convex (in general, the convex hull of the Aumann expectation).

(2) The Vorob'ev expectation of X (Vorob'ev [24], 1984).

(3) The Herer expectation of X (Herer [8], 1990).

(4) The radius-vector expectation of X (Stoyan–Stoyan [18], 1994).

(5) The Vorob'ev median of X (Stoyan–Stoyan [18], 1994).

Another notion covered by our framework is the *coverage function* of a random set X, given by $p_X(x) = P(x \in X)$.

Proposition 8. *Let X be a random closed set. Then, p_X is both an integral depth function in the sense of [22] and a fuzzy set of central points, corresponding to taking $\mathcal{A}$ to be the family of all crisp points and $\mathcal{P}$ the family of all selections of X.*

The corresponding central region is formed by all the *fixed points* of the random set.

7 Choquet capacities

Let ν be a *Choquet capacity* in the sense of [10], namely a function from the Borel σ-algebra of $\mathbf{R}^d$ to $[0,1]$ with the following properties:

a) $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$, $\nu(A) \leq \nu(B)$ if $A \subset B$,

b) $\nu(C_n) \searrow \nu(C)$ if $C_n \searrow C$ and C_n, C are closed,

c) $\nu(A_n) \nearrow \nu(A)$ if $A_n \nearrow A$.

A more general definition, in which closed sets are replaced by compact sets, is possible. A capacity is called *2-alternating* if

$$\nu(A \cup B) + \nu(A \cap B) \leq \nu(A) + \nu(B),$$

and *infinitely alternating* if

$$\nu(\bigcap_{i=1}^{n} A_i) \leq \sum_{\emptyset \neq I \subset \{1,\ldots,n\}} (-1)^{\mathrm{card}(I)+1} \nu(\bigcup_{i \in I} A_i)$$

whenever $A_1, \ldots, A_n$ are Borel sets.

The dual capacity to ν is given by $\overline{\nu}(A) = 1 - \nu(A^c)$. The dual to a 2-alternating or infinitely alternating capacity has the property called 2-monotonicity or infinite monotonicity, respectively. The *Choquet integral* of a random variable ξ with respect to ν is

$$\int_{(C)} \xi \mathrm{d}\nu = \int_0^{\infty} \nu(\{\xi \geq t\})\mathrm{d}t + \int_{-\infty}^{0} [\nu(\{\xi \geq t\}) - 1]\mathrm{d}t,$$

which exists provided at least one of the two improper Riemann integrals is finite. In that case, ξ is called ν-integrable.

Our notion of centrality can be applied to capacities by taking $\mathcal{P}$ to be the core of ν, namely the family of all probability distributions $P \leq \nu$ dominated by ν. Observe that Zadeh's definition of the probability of a fuzzy event extends immediately to capacities by setting $\nu(A) = \int_{(C)} A \mathrm{d}\nu$.

Proposition 9. *Let ν be a 2-alternating Choquet capacity and let I be Goguen's fuzzy implication. Then, taking $\mathcal{P}$ to be the core of ν results in*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), \nu(A)).$$

The Choquet integrals with respect to ν and $\overline{\nu}$ are sometimes called the upper and lower Choquet integrals. They can be retrieved in our framework.

Theorem 10. *Let ν be an infinitely alternating Choquet capacity and $\mathcal{P}$ its core. Let ξ be a ν-integrable random variable. Then, for an appropriate choice of $\mathcal{A}$, the central region C_1 is the interval between $\int_{(C)} \xi \mathrm{d}\overline{\nu}$ and $\int_{(C)} \xi \mathrm{d}\nu$.*

As a consequence, we obtain the following geometric characterizations of the Choquet integral.

Proposition 11. *Let ν be a Choquet capacity, and let ξ be a ν-integrable random variable. If ν is infinitely alternating, then the Choquet integral $\int_{(C)} \xi \mathrm{d}\nu$ is the largest value x satisfying either of the following equivalent properties:*

(1) $|a - x| \leq \int_{(C)} |a - \xi| \mathrm{d}\nu$ for all $a \in \mathbf{R}$,

(2) $d(a, x) \leq \int_{(C)} d(a, x) \mathrm{d}\nu$ for all $a \in \mathbf{R}$.

Analogously, if ν is infinitely monotone, then $\int_{(C)} \xi \mathrm{d}\nu$ is the smallest value x satisfying either property.

8 Concluding remarks

1. The proofs of the results presented in this contribution use pieces of information from several submitted papers, mostly [22, 20] but also [19, 21], which are not explicitly about fuzzy sets. The proofs will be presented in a forthcoming paper. The communication format allows us to present the ideas in a convenient sequence from particular to general, although the logical path of their derivation is more involved.

2. There is an interesting connection between defuzzification and depth-based statistical estimation. Indeed, defuzzification (understood as obtaining a single point, not a crisp set) of a fuzzy set of central points appears very close to location estimation. The two main methods of depth-based estimation in the literature are generalized medians and the depth-weighted estimators given by

$$DWE = \int x D(x) \mathrm{d}x / \int D(x) \mathrm{d}x,$$

see e.g. [27]), which correspond to the maximum and centroid methods of defuzzification.

3. The choice of the Goguen implication, rather than any other fuzzy implication, is rooted in earlier works which inspired our own. The genealogy is as follows. Koshevoy and Mosler [11] defined the *zonoid trimmed regions*

$$\{E[\xi \cdot g(\xi)] \mid g : \mathbf{R}^d \to [0, \alpha^{-1}] \text{ is measurable}\}$$

formed by the expectations of ξ with respect to all probability measures Q such that $\frac{\mathrm{d}Q}{\mathrm{d}P} \leq \alpha^{-1}$. Cascos generalized that idea to an abstract definition of *integral trimmed regions*

$$D^{\alpha}_{\mathcal{F}}(x) = \{x \mid \exists Q \leq \alpha^{-1} P \ : \ \forall f \in \mathcal{F} \ f(x) \leq \int f \mathrm{d}Q\}$$

where $\mathcal{F}$ is a family of integrable functions) which were studied in his Ph.D. thesis and the paper [4]. One already finds $P \geq \alpha Q$ explicitly there, although as a comparison between probabilities. Integral trimmed regions in the sense of [4] receive a natural interpretation in connection to stochastic orders. They coincide with integral trimmed regions in the sense of [22] (no longer interpretable from stochastic orders) if $\mathcal{F}$ is formed by upper semicontinuous functions. In turn, those are a special case of integral central regions [22], which are very similar to central regions in this paper (although, the notions of integral depth in [22] and fuzzy centrality in this paper seem to be more far apart).

4. It is natural to ask whether fuzzy centrality and/or statistical depth are in turn related to Zadeh's proposal that probability theory should include a notion of the *usual value* of a variable.

Acknowledgment

Research has been partially carried out while the author was with the *Universidad de Zaragoza*, and partially funded by Spain's *Ministerio de Educación y Ciencia* (MTM2005-02254 and TSI2005-02511), *Ministerio de Ciencia e Innovación* (MTM2008-01519) and Aragón's *Gobierno* (PM2007-034).

References

[1] R. J. Aumann (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.* 12:1–12.

[2] V. Barnett (1976). The ordering of multivariate data. *J. Royal Statist. Soc. A* 139:318–352 (with discussion).

[3] I. Cascos (2007). The expected convex hull trimmed regions of a sample. *Comput. Statist.* 22:557–569.

[4] I. Cascos, M. López-Díaz (2005). Integral trimmed regions. *J. Multivar. Anal.* 96:404–424.

[5] I. Cascos-Fernández, M. López-Díaz (2004). The zonoid hull of a multivariate probability distribution and its trimmed regions. In: M. López-Díaz, M. Á. Gil, P. Grzegorzewski, O. Hryniewicz, J. Lawry, editors. *Soft methodology and random information systems*, 423–430. Springer, Berlin.

[6] S. Doss (1949). Sur la moyenne d'un élément aléatoire dans un espace distancié [*On the mean of a random element in a metric space*, in French]. *Bull. Sci. Math.* 73:1–26.

[7] P. Hájek, L. Godo, F. Esteva (1995). Fuzzy logic and probability. *Proc. 11th Annual Conf. Artif. Intelligence*, 237–244.

[8] W. Herer (1990). Mathematical expectation and martingales of random subsets of a metric space. *Probab. Math. Statist.* 11:291–304.

[9] P. J. Huber (1972). Robust statistics: a review. *Ann. Math. Statist.* 43:1041–1067.

[10] P. J. Huber, V. Strassen (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* 1:251–263.

[11] G. Koshevoy, K. Mosler (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* 25:1998–2017.

[12] G. Koshevoy, K. Mosler (1998). Lift zonoids, random convex hulls and the variability of random vectors. *Bernoulli* 4:377–399.

[13] H. Kudō (1954). Dependent experiments and sufficient statistics. *Nat. Sci. Rep. Ochanomizu Univ.* 4:151–163.

[14] R. Liu, R. Serfling, D. Souvaine, editors (2006). *Data depth: robust multivariate analysis, computational geometry and applications*. American Mathematical Society, Providence.

[15] R. Y. Liu (1990). On a notion of data depth based on random simplices. *Ann. Statist.* 18:405–414.

[16] R. Y. Liu, K. Singh (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* 88:252–260.

[17] K. Mosler (2002). *Multivariate dispersion, central regions and depth. The lift zonoid approach*. Lect. Notes Statist. 165. Springer, Berlin.

[18] D. Stoyan, H. Stoyan (1994). *Fractals, random shapes and point fields*. Wiley, Chichester.

[19] P. Terán. On the equivalence of Aumann and Herer expectations of random sets. *Test* 17:505–514.

[20] P. Terán. Classical location estimators for random sets in the light of statistical depth. *Submitted for publication*.

[21] P. Terán. Herer and Aumann expectations of random sets. *Submitted for publication*.

[22] P. Terán. Integral central regions and integral depth. *Submitted for publication*.

[23] J. W. Tukey (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, 523–531. Canad. Math. Congress, Montreal.

[24] O. Y. Vorob'ev (1984). *Mean-measure modelling* [in Russian]. Nauka, Moscow.

[25] L. A. Zadeh (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.* 23:421–427.

[26] Y. Zuo, R. Serfling (2000). General notions of statistical depth function. *Ann. Statist.* 28:461–482.

[27] Y. Zuo, H. Cui, X. He (2004). On the Stahel-Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.* 32:167–188.