Discriminant Analysis for fuzzy random variables based on nonparametric regression

Ana Colubi, M.A. Gil, [1] Gil González-Rodríguez, Wolfgang Trutschnig [2]

1.Department of Statistics and OR, University of Oviedo
Oviedo, Spain
2.Unit of Intelligent Data Analysis and Graphical Models, European Centre for Soft Computing
Mieres, Spain
Email: {colubi, magil}@uniovi.es, {gil.gonzalez, wolfgang.trutschnig}@softcomputing.es

Abstract— *This communication is concerned with the problem of supervised classification of fuzzy data obtained from a random experiment. The data generation process is modelled through fuzzy random variables which, from a formal point of view, can be identified with a kind of functional random element. We propose to adapt one of the most versatile discriminant approaches in the context of functional data analysis to the specific case we handle. The discriminant analysis is based on the kernel estimation of the nonparametric regression. The results are applied to an experiment concerning fuzzy perceptions and linguistic labels*

Keywords— fuzzy data, random experiments, supervised classification, kernel estimation, nonparametric regression.

1 Introduction

In many random experiments, as for instance sociological surveys, ecological studies, etc., some characteristics of interest can be assessed in a more meaningful scale if the respondents or the experts are allowed to indicate the degree of precision/imprecision of their answers/judgment/perceptions by means of fuzzy sets (see, for instance, [3] and Section 4 for more details). The results of those experiments can be soundly modelled by means of fuzzy random variables in Puri & Ralescu's sense [16]. It should be remarked that fuzzy random variables in this sense are used to model fuzzy-set-valued random elements and that our goal regards the fuzzy random variables 'per-se'. However, there are other experiments in which the quantity of interest is an underlying real-valued random variables that cannot be precisely observed. In this cases, different approaches may be considered (see, for instance, [5]).

From a formal point of view fuzzy random variables can be identified with a special case of functional random variables, although with some particular features concerning the natural arithmetic and metric structure (see [7] for a deeper discussion). Functional Data Analysis has become an important area of research in recent years (see, for instance, [12, 17, 8, 19]), and as it was suggested in [7], it is possible to take advantage of some of the results developed for those elements to analyze fuzzy data.

As a first step in classification problems concerning fuzzy random variables, some unsupervised approaches have been considered in the literature (see, for instance, [10]). In this communication we deal with the supervised classification problem. That is, given a set of possible groups and a training random sample of fuzzy data of each group, the goal is to predict the group membership of a new fuzzy datum.

As it is indicated in [6], there are different approaches to this problem in the functional context. Most of them are based on modifications of the linear discriminant analysis, however to avoid the inconveniences that frequently arises from the presence of nonlinear class boundaries, we propose a nonparametric method inspired by [6].

The rest of the paper is organized as follows. In Section 2 we introduce the notation and the basic concepts. In Section 3 the classification approach is discussed. Section 4 is devoted to the empirical results, and finally in Section 5 we conclude with some remarks and open problems.

2 Preliminaries

We will denote by $\mathcal{F}_c(\mathbb{R}^p)$ the *class of fuzzy sets* $U : \mathbb{R}^p \to [0,1]$ whose α-levels U_α are nonempty compact convex subsets of $\mathbb{R}^p$ for all $\alpha \in [0,1]$, where $U_\alpha = \{x \in \mathbb{R}^p \,|\, U(x) \geq \alpha\}$ for all $\alpha \in (0,1]$, and $U_0 = \text{cl}(\{x \in \mathbb{R}^p \,|\, U(x) > 0\})$.

Recently, a new class of metrics based on the generalization of the mid-point and the spread of an interval has been defined. This class is very intuitive and versatile and has good properties for the statistical analysis (see [18]).

The generalized mid-point and spread of a fuzzy set A is a way of identifying levelwise the center (location) and the extent (imprecision) by considering each direction in the multidimensional case through the unit sphere $\mathbb{S}^{p-1}$.

Formally, let $\alpha \in [0,1]$ and $u \in \mathbb{S}^{p-1}$, and calculate the lengths $\pi_u(A_\alpha)$ of all orthogonal projections of A_α on this direction, i.e.

$$\pi_u(A_\alpha) = \left[\underline{\pi}_u(A_\alpha)\,,\, \overline{\pi}_u(A_\alpha)\right] = \left[\,-s_{A_\alpha}(-u)\,,\, s_{A_\alpha}(u)\right]$$

where s stands for the support function of a nonempty convex compact set, that is, $s_{A_\alpha}(u) = \sup_{a\in A_\alpha}\langle u,a\rangle$ ($\langle\cdot,\cdot\rangle$ denoting the usual inner product in $\mathbb{R}^p$). Thus, the generalized mid-point and spread of A are defined as the functions $\text{mid}_A, \text{spr}_A : \mathbb{S}^{p-1} \times [0,1] \to \mathbb{R}$ so that

$$\text{mid}_A(u,\alpha) = \text{mid}_{A_\alpha}(u) = \frac{1}{2}\big(s_{A_\alpha}(u) - s_{A_\alpha}(-u)\big),$$

$$\text{spr}_A(u,\alpha) = \text{spr}_{A_\alpha}(u) = \frac{1}{2}\big(s_{A_\alpha}(u) + s_{A_\alpha}(-u)\big).$$

Note that in the interval case, $\mathbb{S}^{p-1} = \{-1,1\}$ and $\text{mid}_A(-1,\alpha) = -\text{mid}\,A$, $\text{mid}_A(1,\alpha) = \text{mid}\,A$, $\text{spr}_A(-1,\alpha) = \text{spr}_A(1,\alpha) = \text{spr}\,A$ holds for all $\alpha \in [0,1]$.

ISBN: 978-989-95079-6-8

The generalized mid-point and spread are not defined as numbers summarizing the 'central tendency' and the 'imprecision', but as functions identifying the 'central points' and the 'imprecision' in the different directions of the Euclidean space. In this way, we get a meaningful characterization of the fuzzy sets alternative to the classical support functions.

The class of distances in [18] is defined from the distances between the level sets as a meaningful generalization of the Bertoluzza et al. metric [1], by considering L_2-type distances between the mid-points and the spreads. Specifically, for each level set $\alpha \in [0,1]$, we define

$$d_\theta^2\big(A_\alpha, B_\alpha\big) = \|\mathrm{mid}\, A_\alpha - \mathrm{mid}\, B_\alpha\|^2 + \theta\|\mathrm{spr}\, A_\alpha - \mathrm{spr}\, B_\alpha\|^2$$

where $\|\cdot\|$ is the usual L_2-norm in the space of the square-integrable functions $L^2(\mathbb{S}^{p-1})$, and $0 < \theta \leq 1$ determines the relative importance of the squared distance between the spreads in relationship with the squared distance between the mids.

The metric between fuzzy sets D_θ^φ is defined as a weighting mean (w.r.t. a probability measure φ with support $[0,1]$) of the distances between the level sets by

$$D_\theta^\varphi(A,B) = \left(\int_{[0,1]} d_\theta^2(A_\alpha, B_\alpha)\, d\varphi(\alpha)\right)^{1/2}$$

The weight measure φ reflects the intuitive (or subjective) interpretation of fuzzy sets - for instance one may treat every α-level as equally important (and therefore use the Lebesgue measure as weight measure φ), or give more mass to α-levels close to 1 or to α-levels close to 0.

If we denote by $\|\cdot\|_{\tilde{2}}$ the usual L_2-norm on the Hilbert space $\mathcal{H} = L^2(\mathbb{S}^{p-1} \times [0,1])$, we have the following alternative intuitive expression for the metric:

$$(D_\theta^\varphi(A,B))^2 = \big\|\mathrm{mid}_{\,A} - \mathrm{mid}_{\,B}\big\|_{\tilde{2}}^2 + \theta\big\|\mathrm{spr}_{\,A} - \mathrm{spr}_{\,B}\big\|_{\tilde{2}}^2.$$

Let $(\Omega, \mathcal{A}, P)$ be a probability space. A *Fuzzy Random Variable* (FRV) can be identified with a Borel measurable mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^p)$ (see [2, 16]). The concepts of induced probability distribution and independence are the same as for general metric-space valued random elements.

3 Nonparametric discriminant approach for fuzzy data

Assume that we have a population $(\Omega, \mathcal{A}, P)$, and for each individual we observe a fuzzy datum. In addition, each individual may belong to one of k different categories $g_1, \ldots, g_k$, and as learning sampling we have the fuzzy data and the group of n independent individuals. The goal is to find a rule that allows us to classify a new individual in one of the k groups from the fuzzy datum. For this purpose we suggest to use a nonparametric approach.

Formally, let $(\mathcal{X}, G) : \Omega \to \mathcal{F}_c(\mathbb{R}^p) \times \{g_1, \ldots, g_k\}$ be a random element in such a way that $\mathcal{X}(\omega)$ is a fuzzy datum and $G(\omega)$ is the membership group (g_1,... or g_k) of each individual $\omega \in \Omega$. Assume that we have n independent copies of $(\mathcal{X}, G)$ as training sample, that is, we have a random sample $\{\mathcal{X}_i, G_i\}_{i=1}^n$. As in a more general functional context (see [6]), we propose to estimate nonparametrically

$$P(G = g_j | \mathcal{X} = \widetilde{x})$$

for $j = 1, \ldots, k$, $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$, and then to assign the new data to the class of higher estimated probability.

In order to find a reasonable estimator of the preceding probability consider first $\delta > 0$ and a closed ball $B(\widetilde{x}; \delta)$ defined in terms of the metric introduced in Section 2. From Bayes Theorem we have that

$$P(G = g_j | \mathcal{X} \in B(\widetilde{x}; \delta))$$

$$= \frac{P(X \in B(\widetilde{x};\delta) | G = g_j) P(G = g_j)}{\sum_{l=1}^k P(X \in B(\widetilde{x};\delta) | G = g_l) P(G = g_l)}$$

$$= \frac{P(D_\theta^\varphi(\mathcal{X}, \widetilde{x}) \leq \delta | G = g_j)) P(G = g_j)}{\sum_{l=1}^k P(D_\theta^\varphi(\mathcal{X}, \widetilde{x}) \leq \delta | G = g_l) P(G = g_l)}$$

For each $\widetilde{x}$, obviously $D_\theta^\varphi(\mathcal{X}, \widetilde{x})$ is a (real-valued) random variable. Let $F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ denote the distribution function of this variable in the group g_j. Consequently,

$$P(G = g_j | \mathcal{X} \in B(\widetilde{x}; \delta))$$

$$= \frac{F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(\delta) P(G = g_j)}{\sum_{l=1}^k F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(\delta) P(G = g_l)}$$

If we assume (as it is natural in the nonparametric setting) that $F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ is uniformly continuous for all $j = 1, \ldots k$, then we have that there exists an underlying density function $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ and

$$\lim_{\delta \to 0} \frac{F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(\delta)}{\delta} = f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$$

As a result,

$$P(G = g_j | \mathcal{X} = \widetilde{x})$$

$$= \frac{f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0) P(G = g_j)}{\sum_{l=1}^k f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(0) P(G = g_l)}.$$

Note that the denominator is just $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$, and for this reason according to [6] we could estimate $P(G = g_j | \mathcal{X} = \widetilde{x})$ by means of

$$\widehat{P}(G = g_j | \mathcal{X} = \widetilde{x})$$

$$\frac{\sum_{i=1}^n I_{G=g_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}{\sum_{i=1}^n K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

$$= \frac{(nh)^{-1} \sum_{i=1}^n I_{G=g_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}{(nh)^{-1} \sum_{i=1}^n K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

ISBN: 978-989-95079-6-8

$$= \frac{\left[(n_j h)^{-1} \sum_{i \in N_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))\right] \cdot (n_j n^{-1})}{(nh)^{-1} \sum_{i=1}^{n} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

where $N_j = \{i \in \{1, \ldots, n\} | G_i = g_j\}$, n_j is the cardinality of N_j, K is a kernel and h the bandwidth.

The last expression shows clearly that the estimator depends on an estimate of $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$ (the densities within each group at the point 0), the empirical estimate of $P(G = g_j)$ and an estimate of $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$ (the overall density at the point 0). In this approach the bandwidth is the same for all the estimates, and it is determined by employing the whole sample. In this way, it is verified that

$$\sum_{j=1}^{k} \widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = 1.$$

However, since that $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$ is composed by a mixture of the variables $\{D_\theta^\varphi(\mathcal{X}, \widetilde{x}) | G = g_j\}_{j=1}^{k}$ whose distributions are expected to be different (recall that the aim is to discriminate among them), it seems more convenient to estimate separately each distribution (each one of them with its own bandwidth) and then to combine them to estimate $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$. Thus, we propose to estimate $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$ by means of

$$\widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0) = (n_j h_j)^{-1} \sum_{i \in N_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}, \widetilde{x}))$$

for all $j = 1, \ldots, k$. In this way,

$$\widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = \frac{n_j n^{-1} \widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)}{\sum_{l=1}^{k} (n_j n^{-1}) \widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(0)},$$

and it is also verified that

$$\sum_{j=1}^{k} \widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = 1.$$

The considered estimators are statistically consistent under the unique assumption that the corresponding densities exist for each $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$ (and the usual regularity conditions are satisfied), however there is no assumption about the changes in the distributions as $\widetilde{x}$ varies.

4 Empirical results

In an experiment about the visual perception of the length of a rule relatively to a maximum, we have shown images like those in Figure 1 to different people. The aim has been to know the relative measure of the dark line (below)in contrast with the light line (above) for these people. They have been allowed to express the lack of precision that they may have by using a fuzzy scale. After a trial with the software, people have seen different rules with random length and they have indicated their perception about the length in the fuzzy scale and the linguistic label (very small, small, medium, large, very large) that they consider suitable for such a length.

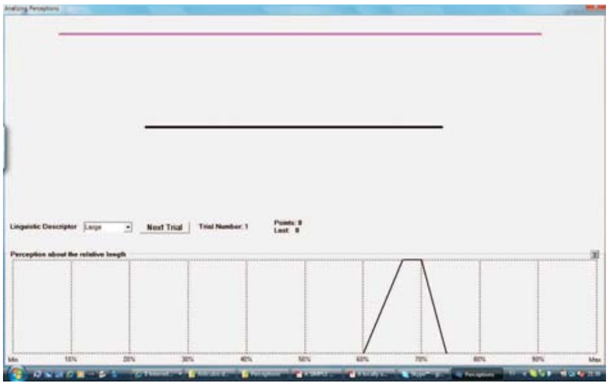

Figure 1: Software to evaluate the visual perception of a rule

The way of using fuzzy scale as explained in the software is as follows:

"*This experiment regards your perception about the relative length of different lines.*

On the top of the screen, we have plotted in light color the longest line that we could show to you. This line will remain visible in the current position during all the experiment, so that you can always have a reference of the maximum length.

At each trial of the experiment we will show you a dark line and you will be asked about its relative length (in comparison with the length of the reference light line):

- *Firstly you will be asked for a linguistic descriptor of the relative length. We have consider five descriptors (Very Small; Small; Medium; Large; Very Large). The aim is to select one of these descriptors at first sign (you can change it later if you want to).*
- *Secondly you will be asked for your own estimate or perception (without physically measuring it) of the relative length (in percentage) by means of a Fuzzy Set (the information about the design and interpretation of the Fuzzy Set will be shown to you at this time).*
- *Finally, in case your initial perception had been changed during the process you can readjust again the linguistic descriptor of its relative length.*"

Concerning the design and interpretation of the fuzzy set, when the image of Figure 1 is shown, the respondents have been asked to draw a fuzzy set representing their perceptions. In accordance with the usual interpretation, the respondents have to choose the 0-level (set of all those points with a positive degree of membership) as the set of all values that they consider compatible with the length to a greater or lesser extent. In the same way, the 1-level (set of all those points with total degree of membership) has to be fixed as the set of values that they consider completely compatible with their perception about the measure of the line. Although it is possible to change the shape of the resulting fuzzy sets, by default the trapezoidal fuzzy set formed by the interpolation of both intervals is fixed (as shown in Figure 1).

In Table 1 we show some of the data of a person who made 220 trials.

The goal here is to predict the category (very small, small, medium, large or very large) that this person would consider

ISBN: 978-989-95079-6-8

Table 1: Perceptions about the relative length of the rule.

Trial	$\inf P_0$	$\inf P_1$	$\sup P_1$	$\sup P_0$	Ling. descrip.
1	78.27	80.94	84.41	87.40	large
2	54.93	58.00	62.20	65.67	large
3	47.25	49.43	50.89	53.31	medium
4	92.65	95.72	97.58	99.11	very large
5	12.92	15.51	17.77	20.03	very small
6	32.55	36.03	39.90	42.89	small
7	2.50	4.44	6.22	9.21	very small
8	24.80	28.19	30.45	33.28	small
9	55.17	58.40	61.79	65.75	large
10	2.26	3.63	5.57	8.08	very small

suitable from the fuzzy perception that he/she has about the length of the line. It should be remarked that the categories are treated here simply as different classes, which may be also labelled as 1, 2, 3, 4 and 5, irrespectively of the fuzzy representation that they may have. The consideration of fuzzy labels would lead to a different approach.

The line showed at each trial has been chosen at random, although to obtain also a validation sample we have proceeded as follows:

- 166 lengths were generated by means of random numbers between 0 and 100. They will be used as training sample.
- The 9 lengths in the equally spaced discrete set $\{100/27 + (i/8)100(1 - 2/27)\}_{i=0,\ldots,8}$ have been repeated 6 times. Thus, we have 54 lengths that are representative of the different situations that may arise, and that can be used as validation sample.
- The 166 random lengths of the training sample and the 54 lengths of the validation sample are interspersed and shown at random.

As it was explained in Section 3, we need to estimate the density of the random variable $D_\theta^\varphi(\mathcal{X}, \widetilde{x})$ at the point 0 for any fuzzy response $\widetilde{x}$ we obtain. The problem of estimating a density in the boundary of its support is not easy. Several approaches can be found in the literature (see for instance, [15], [13] and [11]), however none of them have solved all the inconveniences that may arise. We have considered different options to evaluate its performance in this case study:

M1: Firsty a kernel density estimation with Quartic Kernel

$$K_1(u) = \frac{15}{16}(1 - u^2)^2 I_{(|u| \leq 1)}.$$

and bandwidth chosen by cross-validation were considered. It is well-known that this estimator is non consistent in the boundaries. In particular the estimation at 0 in this case is not good in general (see for instance, [15]). As we have mentioned, several boundary corrections to overcome this problem are available. Many of these corrections (basically based on Jacknife) lead to consistent estimators at the boundaries, although taking negative values in many practical situations (which makes them useless in our case). There are some exceptions in the literature that yield to both consistent and non-negative estimates.

M2: In particular, as a second scenario we have considered the proposal in [6] of using an asymmetric kernel. To be precise, we consider the absolute value of the Quartic kernel K_1, that is,

$$K_2(u) = \frac{15}{8}(1 - u^2)^2 I_{(u \in [0,1])}.$$

This approach leads to consistent estimates at the boundary (but with a low convergence speed ratio). There are no studies about the optimal way of choosing the bandwidth for this asymmetric kernel, so we have used the common one provided by cross-validation techniques applied to this particular case.

M3: Finally, as a third strategy we have considered the estimator proposed by [14] which is a kernel density estimation with boundary correction that provides with non-negative estimates, that is,

$$f_P(0) = \overline{f}(0) \exp\left\{ \frac{\hat{f}(0)}{\overline{f}(0)} - 1 \right\}$$

where $\hat{f}(0)$ denotes the basic kernel density estimator divided by $a_0(0) = \int_{-1}^{0} K(u)du$ and $\overline{f}$ is the boundary corrected kernel density estimator based on generalised jackknifing by combining $\hat{f}$ with the analogous estimator based on the kernel $L(u) = 0.5K(2u)$. In this case and optimal (local) bandwidth is the solution of a complex variational problem (see [15] for a similar situation), that in general is not possible to obtain. For this reason we have considered the global cross-validation bandwidth for the kernel $K = K_1$.

In Table 2 we show the percentage of right classification corresponding to each one of the considered methods. We observe a better behaviour for **M2** than for **M3**. In addition **M1** is quite similar to **M3** in spite of being theoretically inconsistent. This seems to indicate that the selection of a right bandwidth is more important than the theoretical accuracy of the considered estimator (**M1** and **M2** are worse from a theoretical point of view than **M3**, provided that the optimal bandwidth is considered).

Table 2: Percentage of right classification with different methods.

	M1	**M2**	**M3**
% of right classification	75.92	83.33	77.77

5 Concluding remarks

This works is a preliminary study concerning the problem of supervised classification of fuzzy random variables. We propose to use a nonparametric approach to better capture nonlinear boundaries. However, other interesting viewpoints may be used (either by extending those in functional data analysis, as the penalized or flexible discriminant analyses, or by being developed ad-hoc for this case). Further theoretical and empirical comparative studies should be developed.

As it is mentioned in Section 4, the nonparametric estimation problem to be solved here is not easy, and there is no

ISBN: 978-989-95079-6-8

general optimal method to do it. Our empirical results seem to indicate the importance of the bandwidth selection. As it is well-known, the optimal bandwidth for the estimation of a density function is not appropriate in general for the estimation of its curvature. In the same way, the optimal bandwidth for the estimation of a density function is non necessarily appropriated for the classification problem. As an open problem, we propose to optimize the bandwidth selection for the classification problem by choosing, for instance, the values maximizing the proportion of right classifications.

On the other hand, it seems also very interesting to consider the case in which the group membership of the training data is imprecise, as it is done in [4] for real learning data.

Acknowledgment

The research in this communication has been partially supported by the Grant from the Spanish Ministry of Education and Science MTM2006-07501. This financial support is gratefully acknowledged.

References

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

[2] Colubi, A., Domínguez-Menchero, J.S., López-Díaz M., Ralescu, D. A. On the formalization of fuzzy random variables, Inform. Sci. 133 (2001) 3–6.

[3] Colubi, A. 2009. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. Fuzzy Sets and Systems. 160 (3), 344-356.

[4] Côme, E., Oukhellou, L., Denœux, T., Aknin, P. 2009 Learning from partially supervised data using mixture models and belief functions. Pattern Recognition 42 (3), pp. 334-348

[5] Denœux, T., Masson, M.-H., Hébert, P.-A. 2005 Nonparametric rank-based statistics and significance tests for fuzzy data Fuzzy Sets and Systems 153 (1), 1-28.

[6] Ferraty, F., Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Comput. Statist. Data Anal.* **44**, 161–173.

[7] Flórez, J.L., Colubi, A., González-Rodríguez, G., Gil, M.A. 2006. Nonparametric regression with fuzzy random variables through the support functions. In Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty (IPMU'2006, Paris), pp. 724-730.

[8] González-Manteiga, W.G., Vieu, P. Statistics for Functional Data. Computational Statistics and Data Analysis 51 (10), pp. 4788-4792. 2007.

[9] González-Rodríguez, G., Colubi, A., Trutschnig, W. 2009. Simulation of fuzzy random variables. Information Sciences. 179 (5), 642–653.

[10] González-Rodríguez, G., Colubi, A., D'Urso, P., Montenegro, M. 2009. Multi-sample test-based clustering for fuzzy random variables. International Journal of Approximate Reasoning.

[11] Hall, P., Park, B.U. 2002. New methods for bias correction at endpoints and boundaries. The Annals of Statistics 30 (5), 14601479.

[12] Hall, P., Müller, H.-G., Wang, J.-L. 2006. Properties of principal component methods for functional and longitudinal data analysis. Annals of Statistics 34 (3), pp. 1493-1517

[13] Jones, M. C., Foster, P. J. 1993. Generalized jackknifing and higher order kernels. J.Nonparametric Statist. 3, 81-94.

[14] Jones, M. C., Foster, P. J. 1996. A simple nonnegative boundary correction method for kernel density estimation. Statistica Sinica 6(1996), 1005-1013.

[15] Müller, H.-G. 1991. Smooth optimum kernel estimators near endpoints. Biometrika 78, 521-530.

[16] Puri, M. L., Ralescu, D. A. 1986. Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422.

[17] Ramsay, J. O., Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies.* Springer Series in Statistics. Springer-Verlag, New York.

[18] Trutschnig, W. González-Rodríguez, G., Colubi, A., Gil, M.A. 2009. A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Submitted.

[19] Zhang, J.-T., Chen, J. Statistical inferences for functional data. Annals of Statistics 35 (3), pp. 1052-1079. 2007