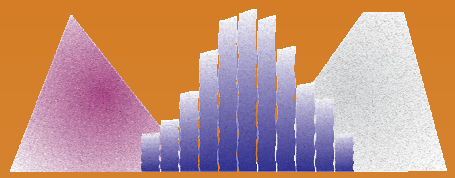
A Proximity-based Discriminant Analysis for Random Fuzzy Sets

Gil González-Rodríguez[1]

Ana Colubi[2], M. Ángeles Gil[2]



SMIRE Research Group (http://bellman.ciencias.uniovi.es/SMIRE)

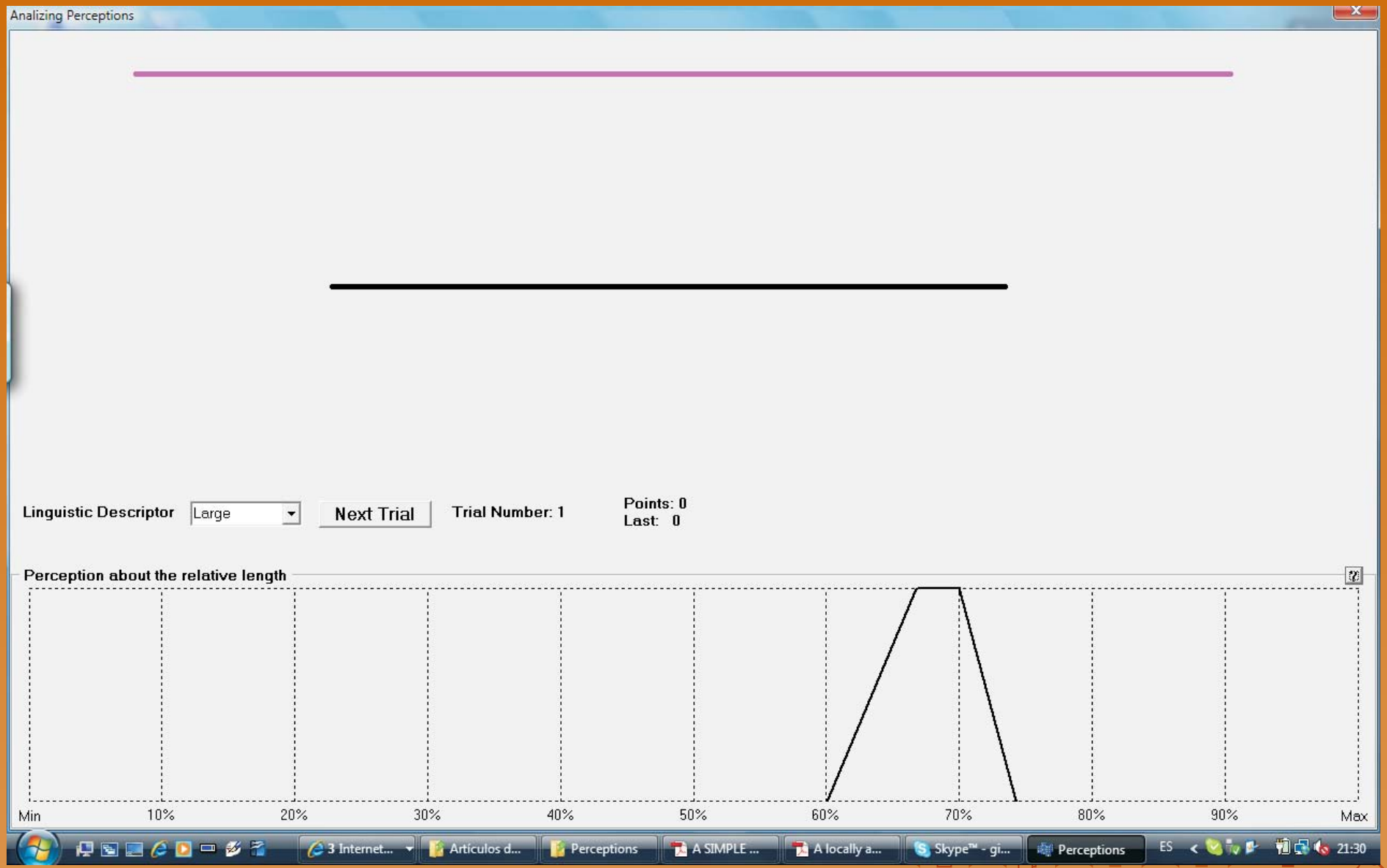
[1]European Centre for Soft Computing, Mieres, Spain

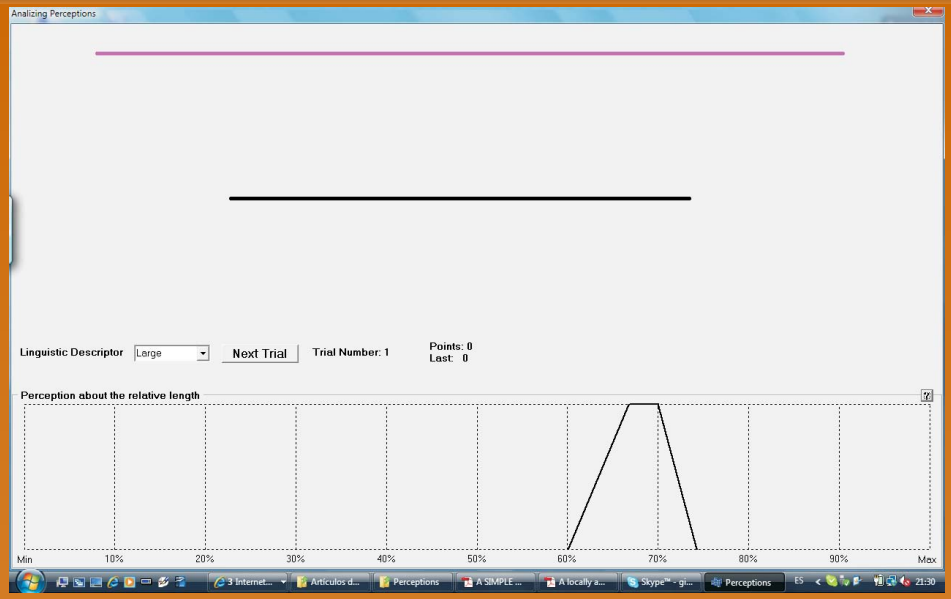
[2]Department of Statistics, Universidad de Oviedo, Spain

COMPSTAT 2010

Paris, August, 2010

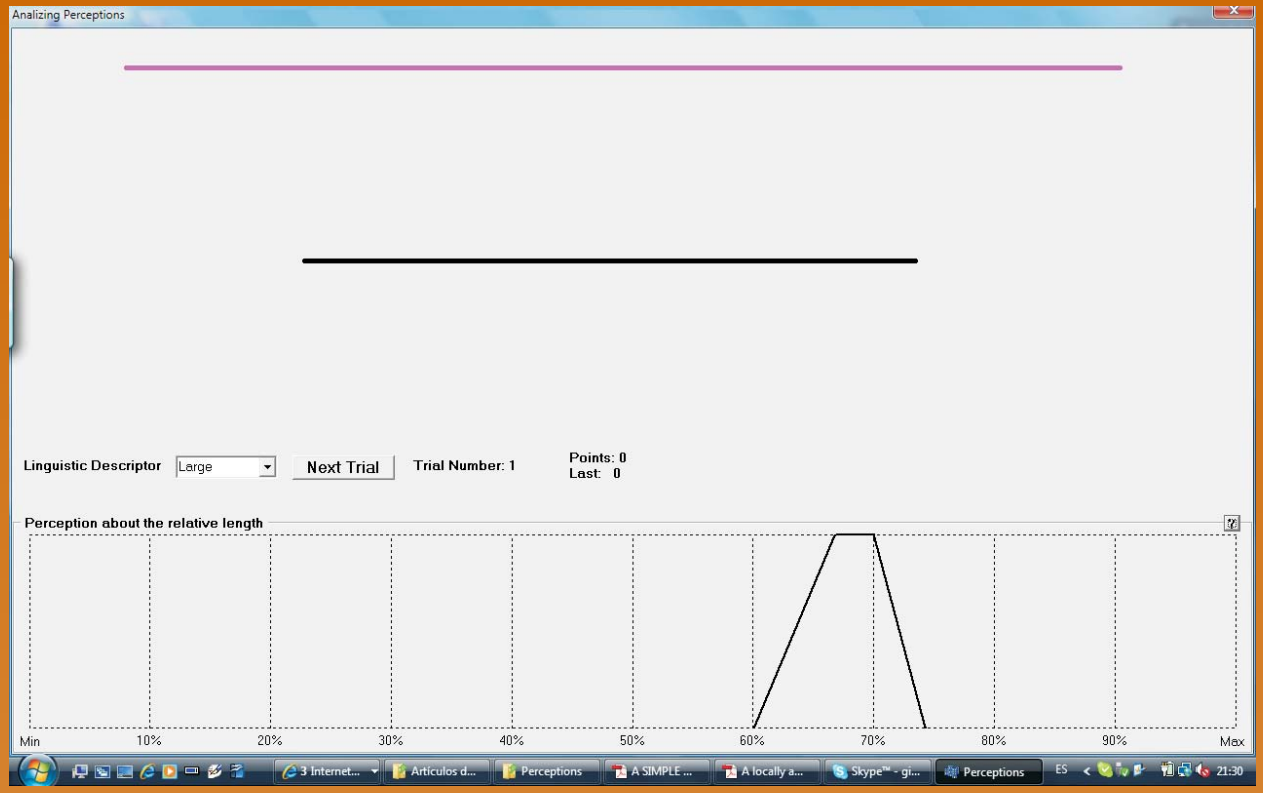
Experiment: perception about the relative length of different lines





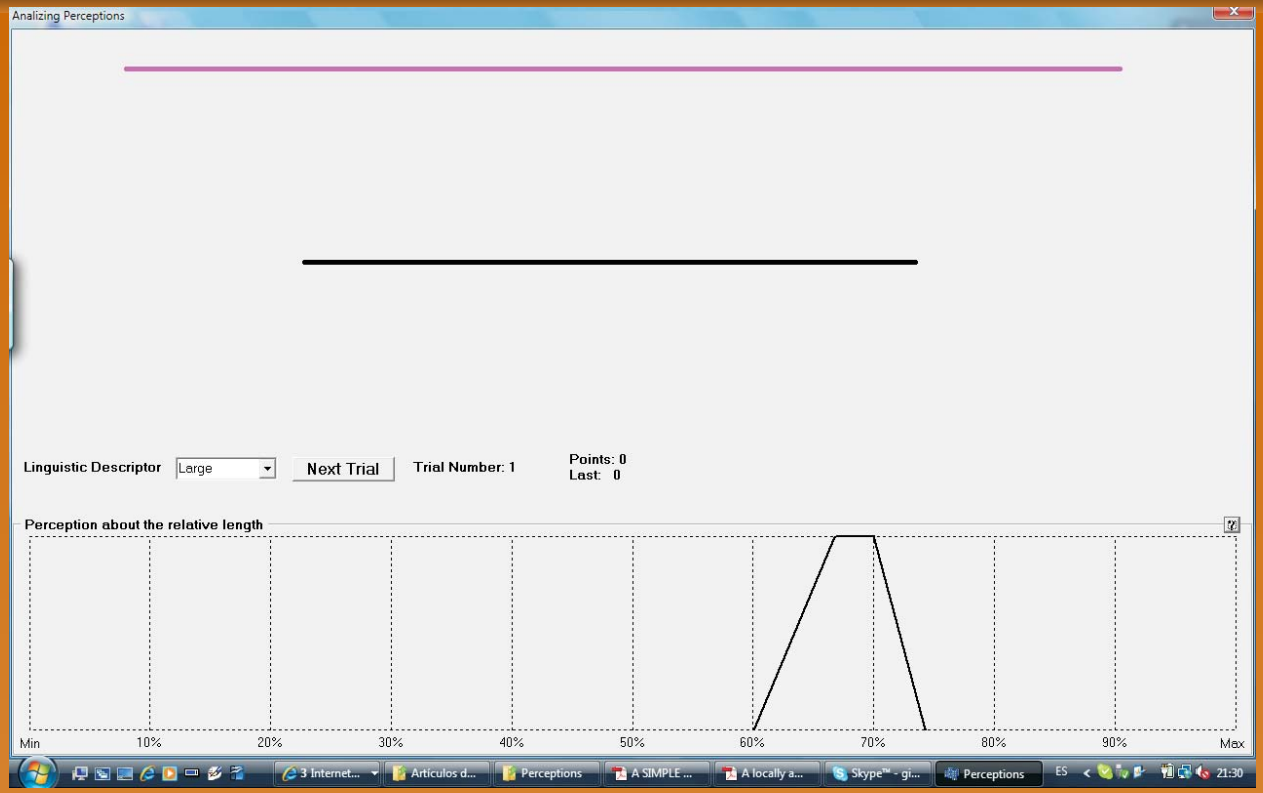
Software explanation: perception about the relative length of lines

- *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)*



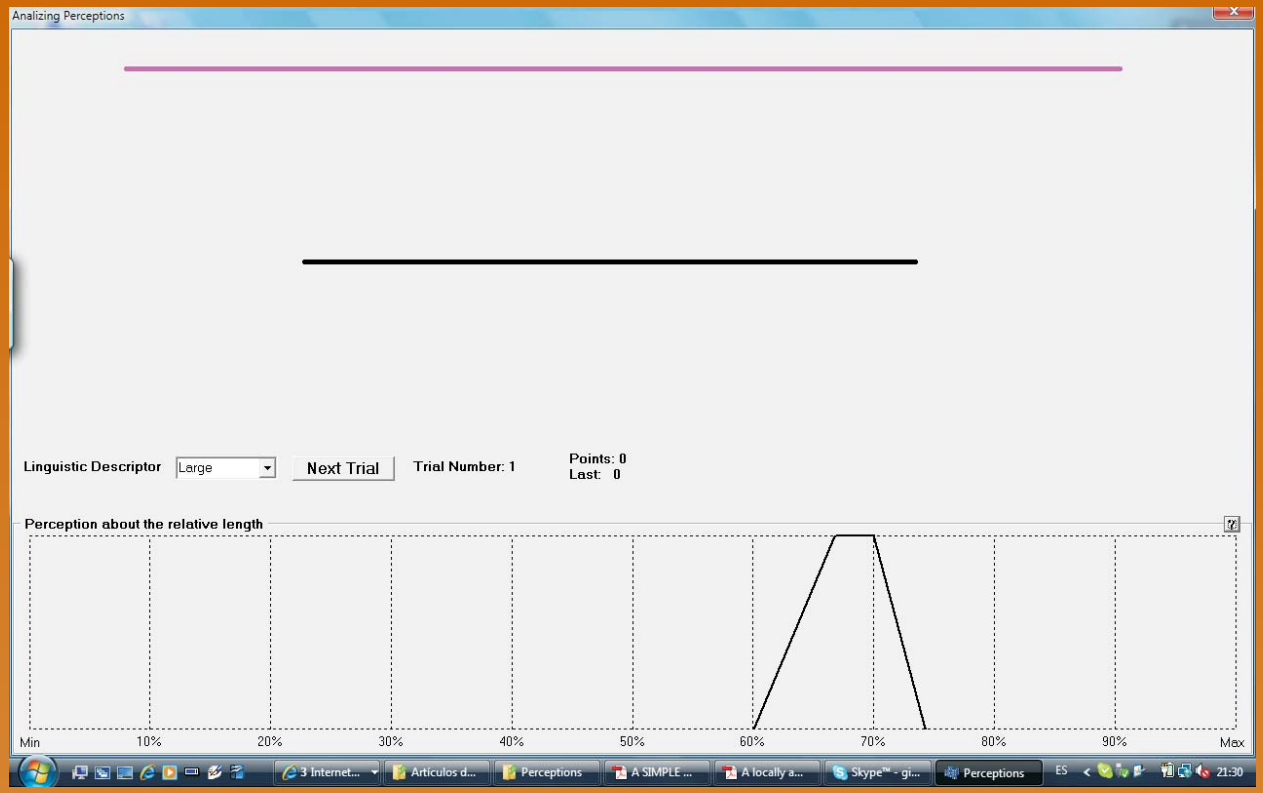
Software explanation

- *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)*



Software explanation

- *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)*

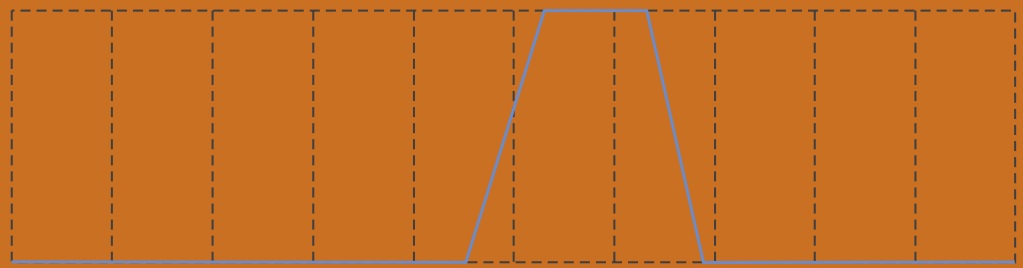


Software explanation

- *Finally, in case your initial perception had been changed during the process you can readjust again the linguistic descriptor of its relative length*

Software explanation: design and interpretation of the fuzzy set

- 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 relative length of the rule to a greater or lesser extent
- 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 length 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



Some data of a person who made 551 trials

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

- http://bellman.ciencias.uniovi.es/SMIRE/perceptions.html

Goal

To predict the category (very small, small, medium, large or very large) that this person would consider suitable from the fuzzy perception that he/she has about the length of the line

- 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

General problem: supervised classification of fuzzy data

- For each individual in a population we observe a fuzzy datum
- Each individual may belong to one of k different categories
- 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
- We suggest to use a Proximity-based Classification Criteria for Fuzzy data approach

The space

- $\mathcal{F}_c(\mathbb{R}^p)$ is the class of fuzzy sets $U : \mathbb{R}^p \to [0,1]$ with nonempty compact convex subsets α-levels U_α
  - ✓ $U_\alpha = \{x \in \mathbb{R}^p \,|\, U(x) \geq \alpha\}$ for all $\alpha \in (0,1]$
  - ✓ $U_0 = \mathrm{cl}(\{x \in \mathbb{R}^p \,|\, U(x) > 0\})$
- From a formal point of view, fuzzy data can be identified with a special case of functional data (with some particular features concerning the natural arithmetic and metric)
- Statistics for fuzzy data can take inspiration from FDA
- L_2 metric based on generalized mid-point and spread
  - ✓ A way of identifying levelwise the center (location) and the extent (imprecision) by considering each direction in the multidimensional case through the unit sphere $\mathbb{S}^{p-1}$

Mid/spread characterization

- Let $\alpha \in [0,1]$, $u \in \mathbb{S}^{p-1}$ and $\langle \cdot, \cdot \rangle$ be the usual inner product in $\mathbb{R}^p$
- s is the support function: $s_{A_\alpha}(u) = \sup_{a \in A_\alpha} \langle u, a \rangle$
- Let $\pi_u(A_\alpha)$ be the set of all orthogonal projections of A_α on this direction, i.e.

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

- Generalized mid-point and spread of A are defined as the functions $\mathrm{mid}_A, \mathrm{spr}_A : \mathbb{S}^{p-1} \times [0,1] \to \mathbb{R}$ so that

$$\mathrm{mid}_A(u,\alpha) = \mathrm{mid}_{A_\alpha}(u) = \frac{1}{2}\left(s_{A_\alpha}(u) - s_{A_\alpha}(-u)\right)$$

$$\mathrm{spr}_A(u,\alpha) = \mathrm{spr}_{A_\alpha}(u) = \frac{1}{2}\left(s_{A_\alpha}(u) + s_{A_\alpha}(-u)\right)$$

The family of distances between $A, B \in \mathcal{F}_c(\mathbb{R}^p)$

- For each level set $\alpha \in [0,1]$

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

  - ✓ $\|\cdot\|$ is the usual L_2-norm in the space of the square-integrable functions $L^2(\mathbb{S}^{p-1})$
  - ✓ $0 < \theta \leq 1$ determines the relative importance of the distances between the spreads w.r.t. the mids

- D_θ^φ is defined as a weighting mean

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

  - ✓ φ is a probability measure with support $[0,1]$ that weights the α-levels as equally important or give more mass to α-levels close to 1 or to α-levels close to 0.

Fuzzy Random Variables (FRVs) and the discriminant problem

- Let $(\Omega, \mathcal{A}, P)$ be a probability space. An FRV can be identified with a Borel measurable mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^p)$
- Let $(\mathcal{X}, G) : \Omega \to \mathcal{F}_c(\mathbb{R}^p) \times \{g_1, \ldots, g_k\}$ be a random element s.t. $\mathcal{X}(\omega)$ is a fuzzy datum and $G(\omega)$ is the membership group $(g_1, \ldots, g_k)$ of each individual $\omega \in \Omega$
- Center of each group: $\mu_j = E(\mathcal{X}|G = g_j)$ $(j \in \{1, \ldots, k\})$
- Relative proximity to each center:
  $R(\widetilde{x}, \mu_j) = P(D_\theta^\varphi(\mathcal{X}, \mu_j) > D_\theta^\varphi(\widetilde{x}, \mu_j)|G = g_j)$
- Training sample: n independent copies of $(\mathcal{X}, G)$, i.e., a random sample $\{\mathcal{X}_i, G_i\}_{i=1}^n$
- Approach: to estimate nonparametrically $R(\widetilde{x}, \mu_j)$ for $j = 1, \ldots, k$, $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$, and then to assign the new data to the class with higher relative proximity

Case-study: details about the design of the experiment

The line showed at each trial has been chosen at random, although to obtain also a good coverage of some interesting situations we have proceeded as follows:

- 479 lengths were generated by means of uniform random numbers between 0 and 100.
- 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.
- All the random lengths were interspersed and shown at random.

Case-study: results

- Percentage of right classification
  - ✓ 10-fold cross validation repeated 100 times

	PCCF	BCCF	$DCCF_1$	$DCCF_2$	$DCCF_3$
(mean)	91.11	90.72	90.41	90.57	88.36
(st.deviation)	0.29	0.22	0.45	0.41	0.53

- PCCF has better mean behaviour than the previous methods in this particular example
- BCCF has the smallest variability
- DCCFs have high variability due mainly to the bandwidth choosen

Concluding remarks

- Preliminary study on a new method for supervised classification of fuzzy random variables
- 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)

Open problems

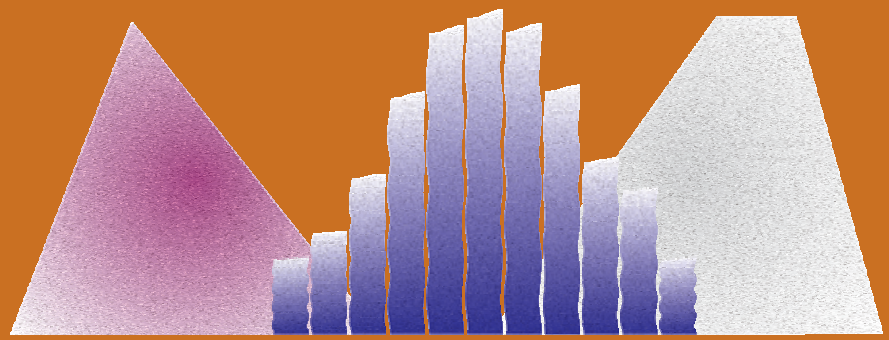
- To develop further theoretical and empirical comparative studies
- To tune the centers by using for instance a weighted mean
- To consider the case in which the group membership of the training data is imprecise

More information...

- Contact
  - ✓ Gil González Rodríguez

    European Centre for Soft Computing

    Mieres. Spain.
    - ✓ gil.gonzalez.rodriguez@gmail.com



Statistical Methods with Imprecise Random Elements