# Advocating the use of Imprecisely Observed Data in Genetic Fuzzy Systems

Luciano Sánchez Computer Science Department Universidad de Oviedo Campus de Viesques 33271 Gijón, Asturias, Spain E-mail: luciano@uniovi.es Inés Couso Statistics Department Universidad de Oviedo Campus de Viesques 33271 Gijón, Asturias, Spain E-mail: couso@pinon.ccu.uniovi.es

*Abstract*— To our opinion, and in accordance with the current literature, the precise contribution of Genetic Fuzzy Systems to the corpus of the machine learning theory has not been clearly stated yet. In particular, we question the existence of a set of problems for which the use of fuzzy rules, in combination with genetic algorithms, produces classifiers inherently better than those arising from the bayesian point of view.

We will show that this set of problems actually exists (and comprises intervalar and fuzzy valued datasets) but it is not being exploited. Current GFSs deal with crisp classification problems, where the role of fuzzy sets is reduced to give a parametric definition of a set of discriminant functions, with a convenient linguistic interpretation. Provided that the customary use of fuzzy sets in statistics is vague data, we propose to test GFSs over imprecisely measured data and design experiments well suited to these problems.

#### I. INTRODUCTION

Statistics and machine learning are closely intertwined. The ambit of application of statistics ranges from experimental designs [20] to theoretical studies about the generalization properties of algorithms, or computational learning theory [17]. Genetic Fuzzy Systems (GFSs), being machine learning tasks, are likewise influenced by this trend, and since recently statistical tests are a standard tool when the performances of genetic classifiers or models are compared [7].

Aside from the machine learning field, fuzzy statistics [2] is an active research area, and there have been advances in the fuzzy counterparts of most of the aforementioned techniques. But, as far as we know, there are scarce connections between fuzzy statistics and GFSs. Contrary to this, we think that the nature of GFSs makes desirable to introduce some elements of fuzzy statistics. In this work we will introduce intervalar and fuzzy extensions of the classification problem, and discuss the impact of these extensions on the design of new Genetic Fuzzy Learning algorithms, and over the experimental design of GFSs.

This work is structured as follows: first, the statistical definition of classification under stochastic noise is introduced, and it is explained that current GFSs are designed to solve the stochastic, crisp problem and not the fuzzy one. Then, an extension to the classification definition is introduced. This extension copes with both stochastic noise and observation

errors, and is based upon the relations between random and fuzzy sets. In section four, it is shown, by means of one example, how the fuzzy valued fitness of the extended GFS is evaluated. The work finishes with the concluding remarks and a discussion about the new opened research lines.

#### II. STATISTICAL CLASSIFIERS AND GFSS

#### A. Classification systems based on discriminant functions

Let us suppose we have a set  $\Omega$  that contains objects  $\omega$ , and let us admit also that each one of them is assigned to a class  $c_i$ ,  $i = 1 \dots N_c$ . We are given a set of measurements  $X(\omega) = (X_1(\omega), \dots, X_{N_i}(\omega))$  over every object. We will say that a classification system is a decision rule that maps every element of  $X(\Omega)$  to a class  $c_i$ , whose main objective is to produce a low number of errors.

For example, let  $\Omega$  be a set of fruits; apples  $(c_1)$ , pears  $(c_2)$  or bananas  $(c_3)$ . We observe the weight and the colour of a randomly selected fruit, for example,  $X(\omega) = (\texttt{yellow}, 150)$ . Our classification system relates the pair (yellow, 150) to the class  $c_3$ , and we wish this relation to be true most of times (i.e., most of the yellow fruits that weight 150 g. are bananas).

Since we do not assume that  $\omega_1 \neq \omega_2 \Rightarrow X(\omega_1) \neq X(\omega_2)$ (i.e., we admit that there can exist a yellow pear weighting 150 g.) perhaps a decision rule that never fails can not be defined for this problem. But an optimum classifier can be defined with respect to the average number of errors.

To define the concept "average number of errors" we need to assume that the mapping X fulfills all necessary conditions to be a random variable. Let us also define a new random variable that quantifies the cost of assigning the class  $c_i$  to an object when it belongs to class  $c_j$ , cost(i, j). If we choose cost(i, j) = 1 when  $i \neq j$  and 0 else, the expectation of the cost function is the mean number of errors. This rule is called "minimum error Bayes rule".

For the problem stated, if the classifier is a decision rule, D(X), and "class( $\omega$ )" is the class of the object  $\omega$ , then the merit value of a classifier can be numerically quantified as

$$\operatorname{err}(D) = \int_{\Omega} \operatorname{cost}(D(X(\omega)), \operatorname{class}(\omega)) \,\mathrm{dP}$$
 (1)

where the error function is integrated with respect to a probability measure P defined over  $\Omega$ . It is well known that this error is optimized by a classifier defined as follows: [16]

$$D(\mathbf{x}) = \arg \max_{i=1,\dots,N_c} P(\text{class}(\omega) = c_i | X = \mathbf{x}).$$
(2)

In practical designs, a monotonic transformation  $M(\cdot)$  of the conditional probabilities  $P(c_i|X)$  does not alter the classification but simplifies computations. We define  $N_c$  functions  $g_i(X) = M(P(c_i|X))$ , taking as the decision rule that X belongs to class *i* for which  $g_i(X)$  is maximum. *g*'s are called "discriminant functions" and the optimal classifier is written

$$D(\mathbf{x}) = \arg \max_{i=1,\dots,N_c} g_i(\mathbf{x}).$$
(3)

### B. GFS should be learnt and evaluated with fuzzy data

It is important to note that this last approach is followed, up to our knowledge, by all Genetic Fuzzy Systems [6]. The random nature of the problem is clearly assumed by all GFS's authors, because current standard experimental designs (leave one out, cross validation, etc.) are unbiased estimations of the classification error over the whole population  $\Omega$  (see eq. 1), and therefore the optimal classifer, no matter the learning technique, is defined by eq. 2, or by one of its transformations, as defined in eq. 3. Moreover, when an input is applied to a fuzzy rule base, the inference process eventually computes  $N_c$  truth values [11] or  $N_c$  number of votes [12] for the set of assertions "the input matches class  $c_i$ " and the defuzzification, in classification problems, consists in chosing the class maximizing the corresponding set of votes. This process is not different from the depicted in eq. 3.

As a consequence of this, the term "fuzzy" does not mean in GFSs that a classification problem different from the crisp one is being solved. "Fuzzy" means here that the parameterising of the discriminant functions has a linguistic interpretation compatible with the fuzzy logic postulates. This does not mean that a fuzzy classifier cannot be fed with fuzzy data; obviously, it can. We mean that neither learning algorithms nor statistical tests take into account the fuzzy nature of the output of the classifier. For example: we know that a random piece of fruit is yellow and weights "about 150 g". Now imagine that we want to compare two classifiers, A and B. Classifier A outputs "pear" with confidence 0.1 and "apple" with confidence 0.2. Classifier B outputs confidences 0.8 and 0.9. Which one is better? Under all statistical tests we are aware of, the two are assigned the same error, because they both eventually will classify the fruit as being an apple.

The experimental designs of GFS that are focused on imprecisely observed data are not being actively studied by the GFS community. Contrary to this, and according to the fuzzy statistics community, the customary use of fuzzy sets in classification and regression problems is the treatment of vague data [10], [15]. We think that this last point should not be understressed. If we admitted that the classification problem being solved by GFSs is not different to the crisp one, it may follow that

- GFSs are not a different machine learning technique than bayesian classifiers. They are a numerical method able to obtain discriminant functions with intuitive meanings.
- There are no reasons different from linguistic interpretability that favour fuzzy rule based classifiers. Therefore, the usefulness of approximate classifiers [5], where linguistic concerns are secondary, is compromised.

By the contrary, if imprecise datasets were used to train and test GFSs, and specific statistical tests were devised to compare classifiers with fuzzy outputs, we would be able to compare different fuzzy classifiers over the set of problems where we expect that fuzzy classifiers make a difference over crisp classifiers, namely datasets with intervalar or fuzzy valued data.

## III. AN EXTENDED DEFINITION OF THE CLASSIFICATION PROBLEM

We will consider that a fuzzy valued dataset is a sample of a fuzzy random variable, as defined in [13], whose  $\alpha$ -cuts are random sets. We will extend first the definition of the classification problem to the intervalar case, and then apply the results to all cuts of the fuzzy random variable sample.

#### A. Intervalar data

Recall eq. 2: to succeed, a learning algorithm should be able to estimate the values  $P(class(\mathbf{x})|\mathbf{x})$  from a sample of measures taken over a subset of  $\Omega$ . To simplify the notation, in this section we assume that X takes values in  $\mathbb{R}$ . When X has absolutely continuous distribution, the standard technique consists in making a transform

$$D(\mathbf{x}) = \arg \max_{i=1,\dots,N_c} \frac{f(\mathbf{x}|c_i)P(c_i)}{f(\mathbf{x})}.$$
 (4)

where f is the density function induced by the random variable X. The denominator can be removed without affecting the result:

$$D(\mathbf{x}) = \arg \max_{i=1,\dots,N_c} f(\mathbf{x}|c_i) P(c_i)$$
(5)

and we obtain a well known result: from an statistical point of view, learning a classifier is the same problem as estimating a density function from a sample of a random variable.

Now we are presented with a sample from a random set, and need to know how can we estimate the density function of the underlying, imprecisely observed random variable (so called *original* random variable in random sets literature [19]). Rephrasing the problem, we need to generalice the concept of density function to the random set case. Our primary thought was to define un "upper" density function as

$$f^*(x) := \lim_{h \downarrow 0} \frac{P^*((x-h,x+h))}{h}$$

provided that this limit exists. For instance, if the random set  $\Gamma : \Omega \to \mathcal{P}(\mathbb{I})$  is a random interval of the form  $\Gamma(\omega) = [X(\omega) - \epsilon, X(\omega)], \forall \omega$ , where X is a random variable with absolutely continuous distribution, this limit exists almost everywhere, but it is  $\infty$ . Observe that  $P^*((x - h, x + h)) =$   $\begin{array}{l} P(\{\omega \in \Omega \mid \Gamma(\omega) \cap (x - h, x + h) \neq \emptyset\}) = P(\{\omega \in \Omega \mid X(\omega) - \epsilon < x + h, X(\omega) > x - h\} = P(\{\omega \in \Omega \mid x - h < X(\omega) < x + h + \epsilon\}). \\ \text{For the continuity of the probability distribution induced by } X, \\ \text{this probability converges to } P_X([x, x + \epsilon]) \\ \text{when } h \\ \text{tends to } 0. \end{array}$ 

To solve this problem, we work directly with eq. 2. We need to estimate the values  $P(c_i|X = \mathbf{x})$ , to choose in each case the *i* for which the corresponding value is maximum (where X represents, in this section, the mentioned original variable, whose imprecise observation is given by  $\Gamma$ .) For each h > 0, we can try to give a couple of upper and lower bounds for the value  $P(c_i|X \in (x - h, x + h))$ . Following [1], the limit when *h* tends to 0 of these quantities is the value we need,  $P(c_i|X = x)$ . Applying the definition of conditional probability, we have that  $P(c_i|X \in (x - h, x + h)) =$ 

$$\frac{P(c_i \cap \{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})}{P(\{\omega \in \Omega \mid X(\omega) \in (x-h, x+h)\})}.$$
(6)

The denominator is, again, the same for all classes, therefore we only need to compare the numerators for the different classes, and give lower and upper bounds for it. The bounds of  $P(c_i \cap \{\omega \in \Omega \mid X(\omega) \in (x - h, x + h)\})$  are

$$\underline{P}_{i}(x,h) = P(c_{i} \cap \{\omega \in \Omega \mid \Gamma(\omega) \subseteq (x-h, x+h)\})$$
(7)

and

$$\overline{P}_i(x,h) = P(c_i \cap \{\omega \in \Omega \mid \Gamma(\omega) \cap (x-h, x+h) \neq \emptyset\}).$$
(8)

Since we do not know the value of  $P_i$ , but a set that contains it, it is clear that, unless the intervals  $[\underline{P}_i, \overline{P}_i]$  do not overlap, we can not know if  $P_i > P_j$  for all pairs of classes, thus the decision rule D is not completely defined. This is is graphically illustrated in Figure 1: the decision rule D is not longer a point function, but a set valued function, where

$$\mathbb{D}(\mathbf{x}) = \{i \mid \nexists j \neq i \text{ with } \underline{P}_i > \overline{P}_i\}$$
(9)

and h is assigned a value small enough for the problem.

Given that ID is a set valued function, the average error of the classifier is not longer known (or, alternatively, we could say that the average error is a set valued statistic.) Anyway, we can find upper and lower bounds for it (see eq. 1). Let us define a pair of functions

$$\underline{\operatorname{cost}}(\omega) = 0 \text{ if } \operatorname{class}(\omega) \in I\!\!D(\Gamma(\omega)), 1 \text{ else}$$
(10)

$$\overline{\operatorname{cost}}(\omega) = 0 \text{ if } \mathbb{I}(\Gamma(\omega)) = \{\operatorname{class}(\omega)\}, 1 \text{ else.}$$
(11)

In words,  $\underline{cost}$  is the optimistic estimation of the error, where we admit that an object is correctly classified if its class number is included in the output, and  $\overline{cost}$  is a pessimistic estimation, where we suppose that an object is misclassified unless its class number is the only output of the classifier. Therefore, the average classification error is contained in the interval

$$\operatorname{err}(I\!\!D) = \left[ \int_{\Omega} \underline{\operatorname{cost}}(\omega) \, \mathrm{dP}, \int_{\Omega} \overline{\operatorname{cost}}(\omega) \, \mathrm{dP} \right]$$
(12)



Fig. 1. Example of three possible situations when comparing interval-valued probabilities. Left: the lower bound of  $P_3$  is higher than the upper bounds of  $P_2$  and  $P_3$ , thus the object is assigned to class 3. Center: the lower bound of  $P_3$  is contained in the range of values of  $P_2$ , but is higher than  $P_1$ , thus the object can be assigned to classes 2 or 3. Right: the intersection of all ranges is not empty, thus the object can be assigned to classes 1, 2 or 3.

#### B. Fuzzy data

If we are given a fuzzy dataset, both the output of the classifier and its expected error will be fuzzy sets, as we show in this section.

Fuzzy datasets can be regarded as samples of a fuzzy random variable  $\widetilde{\Gamma}$ . Every instance of the variable combines two types of noise: random noise, originated in the selection of the object ("we choose a piece of fruit at random") and observation error, originated in an imprecise measure ("the weight of the fruit is high, where 'high' is one of the values of the linguistic variable 'weight").

 $\alpha$ -cuts of  $\widetilde{\Gamma}$  are random sets (for example, the 0.5-cut of the value 'high' can be the interval [100,160]). Therefore, for every value of  $\alpha$  we can build an intervalar classifier, as shown in the preceding section, whose output is a discrete set of class labels ("if the weight is [100,160], then the object is compatible with both pear and apple"). It is intuitive to conclude that the output of the classifier, if presented a fuzzy input, will be a fuzzy set defined over the set of class labels ("if the weight is high, then the object is 0.1/apple+0.6/pear".) The same can be said about the average error of the fuzzy classifier; it will be a fuzzy set.

To obtain this last value, it suffices to admit that the best description we can make about the probability  $P(c_i \cap \{\omega \in \Omega \mid X(\omega) \in (x - h, x + h)\})$ , given that the original random variable X is contained in the fuzzy random variable  $\widetilde{\Gamma}$ , is a fuzzy set  $\widetilde{P}_i$ , whose  $\alpha$ -cuts are intervals  $[\underline{P}_i^{\alpha}, \overline{P}_i^{\alpha}]$  defined as follows:

$$\underline{P}_{i}^{\alpha} = P(c_{i} \cap \{\omega \in \Omega \mid [\widetilde{\Gamma}(\omega)]_{\alpha} \subseteq (x - h, x + h)\}) \quad (13)$$

$$\overline{P}_{i}^{\alpha} = P(c_{i} \cap \{\omega \in \Omega \mid [\widetilde{\Gamma}(\omega)]_{\alpha} \cap (x-h, x+h) \neq \emptyset\}).$$
(14)

Therefore, the fuzzy output of the classifier will be the set

$$[\widetilde{\mathbb{D}}(\mathbf{x})]_{\alpha} = \{i \mid \nexists j \neq i \text{ with } [\underline{P}_j]^{\alpha} > [\overline{P}_i]^{\alpha}\}$$
(15)

and its average error is another fuzzy set,

$$[\widetilde{\operatorname{err}}(I\!\!D)]_{\alpha} = \left[ \int_{\Omega} \underline{\operatorname{cost}}_{\alpha}(\omega) \, \mathrm{dP}, \int_{\Omega} \overline{\operatorname{cost}}_{\alpha}(\omega) \, \mathrm{dP} \right]$$
(16)

where

$$\underline{\operatorname{cost}}_{\alpha}(\omega) = 0 \text{ if } \operatorname{class}(\omega) \in [I\!\!D(\Gamma(\omega))]_{\alpha}, 1 \text{ else}$$
(17)

$$\overline{\operatorname{cost}}_{\alpha}(\omega) = 0 \text{ if } [I\!D(\Gamma(\omega))]_{\alpha} = \{\operatorname{class}(\omega)\}, 1 \text{ else.}$$
 (18)

#### C. Computer-friendly definition

In the preceding subsection we have stated that the average error of a classifier, when its input comprises fuzzy sets, should also be a fuzzy set. Therefore, the fitness functions in GFSs will return a fuzzy value. This value can be numerically estimated by means of eqs. 16, 17 and 18. Since these equations are expressed in terms of a family of  $\alpha$ -cuts, we give a rewriting of them that is easier to codify in a computer.

Let  $\widetilde{ID} = \sum_{i=1...N_c} \mu_i/i$  be the fuzzy output of the classifier, p the class label with maximum membership value in  $\widetilde{ID}$ ,  $p = \arg \max_{i=1...N_c}(\mu_i)$ , and q the second maximum membership,  $q = \arg \max_{i=1...N_c, i \neq p}(\mu_i)$ . Let the height of  $\widetilde{ID}$ ,  $\mu_p = 1$ . Then, the contribution of the object  $\omega$  to the total error is

$$\widetilde{err}(\omega) = \begin{cases} 1/0 + \mu_q/1 & \text{if } \text{class}(\omega) = p\\ \mu_{\text{class}(\omega)}/0 + 1/1 & \text{otherwise.} \end{cases}$$
(19)

For example, suppose that, in a problem with three classes, the output of the classifier is the fuzzy set  $\{0.2/\text{pear}+1/\text{apple}+0.8/\text{banana}\}$ . If the object was a pear, the accumulated error of the classifier would be increased by the fuzzy amount  $\{0.2/0+1/1\}$ . If it was an apple, the new error would be  $\{1/0+0.8/1\}$  higher, or  $\{0.8/0+1/1\}$  if it was a banana.

#### IV. EXAMPLE OF FITNESS EVALUATION IN THE EXTENDED CLASSIFIER

In this section we will numerically evaluate a fuzzy classifier over a small problem, to illustrate the ideas introduced in the preceding section.

Let us suppose that we have to discriminate between three classes (apple, pear, banana), given the weight of a piece of fruit. To design the classifier, we are given a sample comprising five pieces, whose weights and classes are given in table I. Weights are triangular fuzzy numbers, designated by three numbers: leftmost, center and rightmost values.

Let us also suppose that the GFS has to evaluate the fitness of the rule base that follows:

if weight is small then banana if weight is medium then apple if weight is high then pear

	crisp weight	fuzzy weight	class
1	111	(102,111,116)	pear
2	96	(88,96,112)	apple
3	116	(104,116,128)	pear
4	91	(83,89,90)	banana
5	101	(91,101,118)	apple

TABLE I Dataset for the example problem 'fruit'



Fig. 2. Definition of the linguistic variable "weight", as used in the example problem "fruit."

where the linguistic variable "weight" takes the values shown in figure 2. We wish to assign a fitness value to this rule base, given the mentioned dataset.

Observe that the fitness value assigned to this rule base measures the classifier error as defined in eq. (16). This, on one hand, assess the degree to which these rules approximate the usual Bayes criterium in eq. (5), but, on the other hand, also takes into account how sensible these rules are to measurement errors. The nonspecificity of the fuzzy fitness value is higher when bases are less robust, thus the ranking of the fitness values should take into account more information than the given, for example, in the COG of the set.

Let us evaluate first this classifier over the crisp dataset given by the column "crisp weight" in table I. The output of the classifier, using max-min inference, is as follows:

Crisp Input	Output	Cost
111	apple	1
96	apple	0
116	apple	1
89	banana	0
101	apple	0

therefore, the cost of this classifier is 2 (in other words, we estimate that it is wrong 40% of times).

If we apply an intervalar input to the same classifier (the support of the fuzzy examples,) its output is a crisp subset of the class labels. For example, the interval [88,112] has associated the crisp subset  $ID(\Gamma(\mathbf{x})) = ID([88,112]) = \{\text{banana, apple}\}$ , because, if we classify all the points in [88, 112], we observe that points in [88, 92.5] are assigned the class "banana", and points in (92.5, 112] are assigned the class "apple." To calculate the cost of the classifier we operate as follows:

Fuzzy Input	Output	Cost
(102, 111, 116)	{ 1/apple }	$\{1/1\}$
(88, 96, 112)	$\{ 0.5625/banana+1/apple \}$	$\{1/0 + 0.5625/1\}$
(104, 116, 128)	{ 0.875/apple+1/pear }	$\{0.875/0 + 1/1\}$
(83, 89, 90)	{ 1/banana }	$\{1/0\}$
(91, 101, 118)	{ 0.15/banana+1/apple+0.0294/pear }	$\{1/0 + 0.15/1\}$

TABLE II

OUTPUT OF THE EXAMPLE CLASSIFIER WHEN THE INPUT IS A FUZZY SET

Interval Input	Output	Cost
[102, 116]	{ apple }	{1}
[88, 112]	{ banana, apple }	$\{0, 1\}$
[104, 128]	{ apple, pear }	$\{0, 1\}$
[83, 90]	{ banana }	$\{0\}$
[91, 118]	{ banana, apple, pear }	$\{0, 1\}$

therefore the cost is contained in the interval [1, 4] (i.e., when data is precisely measured, we estimated that the classification was wrong 40% of times; when data is intervalar, all we can say without assuming a random distribution of the observation error is that it is wrong between 20% and 80% of times.)

Finally, if the classifier is applied a fuzzy input, its outputs and costs are shown in table II. The inputs are fuzzy triangular numbers, and the data (x, y, z) are left, center and right point. The cost of the classifier is  $\{1/1\} \oplus \{1/0 + 0.5625/1\} \oplus$  $\{0.875/0+1/1\} \oplus \{1/0\} \oplus \{1/0+0.15/1\} = \{0.875/1+1/2+$ 0.5625/3 + 0.15/4, or in words, the error of the classifier is 20% with confidence 0.875, 40% with confidence 1, 60% with confidence 0.5625 and 80% with confidence 0.15. The error still is between 20% and 80%, but a genetic algorithm could prefer this result over a different classifier that has, say  $\{0.875/1 + 1/2 + 0.5625/3 + 0.25/4\}$ , even in the punctual estimations of the classification error of either are the same, because the differences in the fuzzy errors state that the former classifier is less afected by imprecision in the input data (it is less likely to obtain an 80% error.) Observe also that, if input data are triangular fuzzy sets, the punctual error of the classifier is given by the value with membership 1 in the fuzzy cost, 2 (or %40 of errors) in this example.

It is remarked that the algorithm used here to calculate the output, and the error of the classifier, given a fuzzy input, does *not* produce the same results that we would have obtained by means of the direct use of fuzzy inference. For example, if we apply max-min inference to compute the output of the rule base when its input is the fuzzy set (91, 101, 118) we obtain  $\{0.4/banana+0.9048/apple+0.3095/pear\}$ . By the contrary, the procedure proposed in this paper produces the set  $\{0.15/banana+1/apple+0.0294/pear\}$ . In other words, we have proposed to use fuzzy logic to assign a class to a crisp input, but a fuzzy statistics-based interpretation of the observation error to extend the classification to imprecise data.

#### V. CONCLUDING REMARKS AND OPEN PROBLEMS

We have suggested in this paper that the kind of problems where GFSs are inherently better than their stochastic counterparts is composed by those problems including imprecisely observed data. To deal with imprecise data, stochastic algorithms have to introduce additional hypotheses, as a probability distribution (uniform, gaussian, etc.) over the measurement errors, that fuzzy algorithms do not need.

When the classification problem is extended to fuzzy data, there are some changes that have to be done in the numerical algorithms. We enumerate these changes and briefly comment on the problems that these modifications leave open:

- Since the error of a classifier will not be a scalar, but a fuzzy set, the fitness function in the extended GFS must return a fuzzy value. Genetic Algorithms must do fitness-based orderings of individuals, therefore we must a) use a fuzzy ranking [3] to induce a total order over fuzzy parts of [0, 1] or b) induce only a partial ordering and use multicriteria genetic algorithms instead [14]. The selection of the best fuzzy ranking, or, from a more general point of view, processing fuzzy values when evaluating the fitness function in genetic algorithms, is a problem that can not yet be considered as generally solved [18].
- 2) New statistical tests have to be designed in order to judge the relevance of the differences of two fuzzy algorithms. There are some works in fuzzy statistical inference [4], [8], [9], but more work needs to be done in order to do practical comparisons between fuzzy valued algorithms. It is not clear yet how the comparison between fuzzy and crisp data should be done (and we need that, to compare extended GFSs to other algorithms,) and this poses other open problems: the definition of statistical contrasts about fuzzy-valued parameters in fuzzy random variables, the very definition of parametric families of random sets or fuzzy random variables, and the design of their corresponding tests.
- 3) The benchmarks most commonly used to compare GFSs include missing values, that can be codified with fuzzy information, and linguistic data, that can also be assimilated to fuzzy sets, but we lack datasets of imprecisely measured data that allow us to compare the robustness of GFSs to that of stochastic methods in terms of the degree of imprecision in the data. This absence prevent us from optimizing GFSs towards the main objective of fuzzy techniques, as stated by Zadeh [21]: "Exploit the tolerance for imprecision [...] to achieve tractability,

robustness, and low solution cost".

#### ACKNOWLEDGMENT

This work was supported in part by the Spanish Ministry of Education and Science, under grants TIC2002-04036-C05-05 and MTM2004-01269.

#### REFERENCES

- [1] R. B. Ash. Basic Probability Theory. Wiley. 1970.
- [2] C. Bertoluzza, M. A. Gil, D.A. Ralescu, Eds. *Statistical Modeling*, *Analysis and Management of Fuzzy Data* Series in Fuzziness and Soft Computing, 87, Springer. 2003
- [3] L. Campos, A. Gonzlez. "Further contributions to the study of the average value for ranking fuzzy numbers" *International Journal of Approximate Reasoning* 10, 2, pp. 135-153, 1994.
  [4] M.R. Casals, M.A. Gil, P. Gil. "On the use of Zadeh?s probabilistic
- [4] M.R. Casals, M.A. Gil, P. Gil. "On the use of Zadeh?s probabilistic definition for testing statistical hypotheses from fuzzy information". *Fuzzy Sets Syst.*, 20, pp. 175?190, 1986.
- [5] O. Cordon, F. Herrera. "A general study on genetic fuzzy systems", in J. Periaux, G. Winter, M. Galan, P. Cuesta, Eds., Genetic Algorithms in Engineering and Computer Science. Wiley. pp. 33-57. 1995
- [6] O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena Genetic Fuzzy Systems, Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. World Scientific, 2001
- [7] O. Cordon, F. A. C. Gomide, F. Herrera, F. Hoffmann, L. Magdalena. "Genetic fuzzy systems. New developments". Fuzzy Sets and Systems 141, 1, 1 pp 1-3. 2004.
- [8] N. Corral and M.A. Gil. "The minimum inaccuracy fuzzy estimation: an extension of the maximum likelihood principle". *Stochastica*, 8, pp. 63?81, 1984.
- [9] M.A. Gil, N. Corral, P. Gil. "The minimum inaccuracy estimates in χ<sup>2</sup> tests for goodness of fit with fuzzy observations". J. Stat. Plan. Inf., 19, pp 95?115, 1985.
- [10] I. Couso, L. Sanchez, P. Gil "Imprecise distribution function associated to a random set" *Information Sciences*, 159, 1-2, pp. 109 - 123, 2004.
- [11] Dubois, D. Fuzzy Sets and Systems, Theory and Applications. Academic Press, 1980.
- [12] Kuncheva, L. I. Fuzzy Classifier Design. Springer-Verlag, NY, 2000.
- [13] F. Feron. "Ensembles aleatoires flous", C.R. Acad. Sci. Paris Ser. A 282 pp. 903-906. 1975.
- [14] Fonseca, C., Fleming, P.J. "An Overview of Evolutionary Algorithms in Multiobjective Optimization," in Evolutionary Computation 3, pp. 1-16, 1995.
- [15] M.A. Gil. "Fuzziness and loss of information in statistical problems". *IEEE Trans. SMC*, 17, pp. 1012-1025, 1987.
- [16] D. J. Hand Discrimination and classification. Wiley. 1981
- [17] R. Herbrich, T. Graepel "Introduction to the Special Issue on Learning
- Theory". *JMLR*, Special Issue on Learning Theory 4, pp. 755–757, 2003.
  [18] Köppen, M., Franke, K. and Nickolay, B. Fuzzy-Pareto-Dominance driven multiobjective genetic algorithm. Proc. 10th International Fuzzy Systems Association World Congress (IFSA), Istanbul, Turkey, pp. 450-453, 2003
- [19] R. Kruse, K.D. Meyer. Statistics with Vague Data Vol. 33. Reidel, Dordrecht, 1987.
- [20] C. Nadeau, Y. Bengio, "Inference for the Generalization Error" Machine Learning, 52, pp. 239-281, 2003.
- [21] Zadeh, L. A. "Soft Computing and Fuzzy Logic". *IEEE Software*. pp. 48-56. 1994.