Identification of fuzzy measures from sample data with genetic algorithms*

Elías F. Combarro (elias@aic.uniovi.es)†
Artificial Intelligence Center, University of Oviedo, Spain

Pedro Miranda (pmiranda@mat.ucm.es)
Department of Statistics and O. R. , Complutense University of Madrid, Spain

Abstract

In this paper we introduce a method for the identification of fuzzy measures from sample data. It is implemented using genetic algorithms and it is flexible enough to allow the use of different subfamilies of fuzzy measures for the learning, as k-additive or p-symmetric measures. The experiments performed to test the algorithm suggest that it is robust in situations where there exists noise in the considered data. We also explore some possibilities for the choice of the initial population, which lead to the study of the extremes of some subfamilies of fuzzy measures, as well as the proposal of a method for random generation of fuzzy measures.

Keywords: Genetic algorithms, fuzzy measures, k-additivity, p-symmetry.

1 Introduction

Fuzzy measures [51] (also called capacities [11] or non-additive measures [16]) constitute a generalization of classical probability distributions in which we have removed additivity and monotonicity is imposed instead. This extension is perfectly justified in many practical situations, in which additivity is too restrictive. For example, in the field of Decision Making, models based on Probability, as those from von Neumann and Morgenstern [53] or Anscombe and Aumann [3] to cite a few, can lead to inconsistencies due to *risk aversion*, as the well-known paradoxes of Ellsberg [20] or Allais [2]. However, models based on fuzzy measures [9, 47] are able to handle and interpret these problems. Moreover, in recent years, the analysis and use of fuzzy measures have been enriched by different equivalent representations of a capacity [26], that are obtained through invertible linear transformations applied on the measure.

Fuzzy measures have been successfully applied to model problems in Multicriteria Decision Making and Cooperative Games. In the former case, fuzzy measures allow the decision maker to introduce vetoes and favors in the model [26], as well as interactions among the different criteria [27]. In the theory of Cooperative Games, fuzzy measures represent the strength of coalitions of players; they are related to the Shapley value [50], as shown in [25]. Other fields related to fuzzy measures are combinatorics [46], pseudo-boolean functions [29], etc. This versatility of fuzzy measures has led to a huge number of related works, both from a theoretical and from a practical point of view [18, 54].

However, despite the fact of the many advantages of fuzzy measures, their practical use has to face with the hurdle of an increment in the complexity. In the case of finite spaces of cardinality n, just $n-1$ values suffice to define a probability measure, while 2^n-2 coefficients are needed for fuzzy measures. This exponential complexity is the *Achilles heel* of the theory. In an attempt to cope with the complexity involved by the use of fuzzy measures, additional constraints on the measure have been imposed, leading to different subfamilies. For example, Grabisch [25] has proposed the concept of k-additive measures, and Miranda and Grabisch [40] have recently introduced a generalization of symmetric measures, the so-called p-symmetric measures. In both

*This work is a revised an extended version of [12], presented at the IFSA 2003 Conference.

†Corresponding author. Postal Address:Despacho 1.1.36 - Edificios Departamentales - Campus de Viesques - Gijón. 33207 - Spain. Telephone: (+34) 985 18 25 14. Fax: (+34) 985 18 21 25

cases, we obtain an important reduction in the number of coefficients needed to define the fuzzy measure. These families provide models which are both flexible and simple to use.

In practice, we usually have information about some prototypical examples and the problem consists in finding the fuzzy measure (possibly, restricted to a subfamily) that best fits the data. We will assume that the sample information is numerical; if this is not the case, it should be transformed into numerical data through a tool dealing with ordinal values, as MACBETH [19] or TOMASO [37]. It must be noted that the examples often have some level of noise and they also might lead to several, equally suitable, solutions [42].

If the considered proximity criterion is the squared error, different techniques exist to solve the problem. For example, Grabisch and Nicolas [28] have developed a method based on solving a quadratic problem, in [8] Beliakov et al. acknowledge the case when the measures are symmetric or 2 or 3-additive, and in [24], Grabisch proposes an *ad hoc* algorithm for the problem. On the other hand, Wang *et al.* [55] have developed a method based on genetic algorithms [23]; although in these two cases a suboptimal solution is found, the computational is, sometimes, greatly reduced.

In this paper we introduce a new method based on genetic algorithms to solve the problem of learning fuzzy measures from sample data. These algorithms have been successfully applied to many optimization and searching problems and their good properties (robustness, ease of implementation, low computational complexity, ...) make them an appealing choice for our purposes. The method proposed is not restricted to a single family of fuzzy measures, but allows to use any convex subfamily of fuzzy measures.

The rest of the paper is organized as follows: In Sections 2 and 3 we briefly introduce the basic concepts on fuzzy measures and genetic algorithms, respectively; in Section 4 we describe our algorithm and show the importance of the initial population. This leads to the problem of finding the extremes of some subfamilies, which is addressed in Section 5. In Section 6 we present some of the experiments we have conducted with the algorithm, and in Section 7 we propose several alternative methods for the choice of the initial population. Finally, in Section 8 we draw some conclusions and propose some lines for future research.

2 Fuzzy measures

Let us define the basic concepts that will be needed throughout the paper:

Definition 1 *[11, 16, 51] A* **fuzzy measure** *or* **non-additive measure** *on a set $X = \{x_1, \ldots, x_n\}$ of n criteria is a function μ that assigns to each subset of X a real value between 0 and 1 satisfying*

1. *$\mu(\emptyset) = 0$ and $\mu(X) = 1$.*
2. *If $A \subseteq B$ then $\mu(A) \leq \mu(B)$.*

These measures represent the degree of importance of each subset of criteria. The interest of fuzzy measures for decision making relies on the fact they are able to deal with interactions between criteria [27], as well as to model veto and favor situations [26].

We will denote the set of all fuzzy measures by $\mathcal{FM}$. Remark that $\mathcal{FM}$ is a bounded convex polyhedron (i.e., the intersection of a finite number of semispaces).

Definition 2 *Given a convex subset of fuzzy measures, $\mathcal{F} \subseteq \mathcal{FM}$, we say that $\mu \in \mathcal{F}$ is a* **vertex** *or* **extreme point** *of $\mathcal{F}$ if it cannot be written as a convex combination of two different measures of $\mathcal{F}$.*

As $\mathcal{FM}$ is a bounded convex polyhedron, any measure μ can be put as a convex combination of the vertices (see [7]). We will have more to say about the extreme points in Section 5.

A special class of fuzzy measures (which are, in fact, the vertices of $\mathcal{FM}$, see Proposition 1 below) is the set of $\{0,1\}$-valued measures.

Definition 3 *A fuzzy measure is* **$\{0,1\}$-valued** *if it only takes values 0 and 1.*

Notice that for a $\{0,1\}$-valued measure μ, there are some subsets A satisfying the following conditions:

$$\begin{array}{ll} \mu(A) = 1, & \\ \mu(B) = 1, & \forall B \supseteq A, \\ \mu(C) = 0, & \forall C \subset A. \end{array} \tag{1}$$

This leads us to introduce the following concept:

Definition 4 *Consider a $\{0,1\}$-valued measure μ. We will say that a subset A of X is a* **μ-minimal subset** *(or just* **minimal subset** *for short) if it satisfies the conditions 1.*

Definition 5 *[44] The* **unanimity game** *over $A \subseteq X$, $A \neq \emptyset$ is a fuzzy measure defined by*

$$u_A(B) := \begin{cases} 1 & \text{if A} \subseteq \text{B} \\ 0 & \text{otherwise} \end{cases}$$

For $\emptyset$, we define the unanimity game by

$$u_\emptyset(B) := \begin{cases} 1 & \text{if B} \neq \emptyset \\ 0 & \text{if B} = \emptyset \end{cases}$$

Indeed, u_A for $A \neq \emptyset$, are those $\{0,1\}$-valued measures which possess just one minimal subset.

Remark 1 *A $\{0,1\}$-valued measure is completely defined by its minimal subsets. To see this, it suffices to remark that $\mu(A) = 1$ if it contains a μ-minimal subset and $\mu(A) = 0$ otherwise. In fact, $\mu = \vee_{B \in M} u_B$ where $M = \{B : B \text{ is } \mu - minimal\}$.*

Notice that $2^n - 2$ coefficients are needed in order to determine a fuzzy measure on n criteria. This fact makes the use of these measures infeasible in practice. To deal with this problem, several sub-families of fuzzy measures have been proposed in the literature, as k-intolerant measures [36], λ-measures [51], ... In this paper we will need k-additive measures [27] and p-symmetric measures [40].

In order to define k-additive measures we need to introduce the concept of Möbius transform.

Definition 6 *[46] Let μ be a set function (not necessarily a fuzzy measure) on X. The* **Möbius transform (or inverse)** *of μ is another set function on X defined by*

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B), \, \forall A \subseteq X.$$

The Möbius transform given, the original set function can be recovered through the *Zeta transform* [10]:

$$\mu(A) = \sum_{B \subseteq A} m(B).$$

The value $m(A)$ represents the strength of subset A in any coalition in which it appears. The Möbius transform corresponds to the *basic probability mass assignment* in Dempster-Shafer theory of evidence [48].

Definition 7 *[25] A fuzzy measure μ is said to be* **k-order additive** *or* **k-additive** *if its Möbius transform vanishes for any $A \subseteq X$ such that $|A| > k$ and there exists at least one subset A of exactly k elements such that $m(A) \neq 0$.*

In this sense, a probability measure is just a 1-additive measure; thus, k-additive measures generalize probability measures, that are very restrictive in many situations; they fill the gap between probability measures and general fuzzy measures. For a k-additive measure, the number of coefficients is reduced to

$$\sum_{i=1}^{k} \binom{n}{i}.$$

The concept of k-additivity has been extended by Mesiar in [38] for infinite referentials without the need of the Möbius transform. More about k-additive measures can be found e.g. in [27]. We will denote the set of all k'-additive measures with $k' \leq k$ by $\mathcal{FM}^k$; we will use the fact that $\mathcal{FM}^k$ is a convex polyhedron (the proof is straightforward considering the Möbius transform). Specially appealing is the 2-additive case, that provides a generalization of probability allowing interactions while keeping a reduced complexity.

Definition 8 *A fuzzy measure is said to be* **symmetric** *if it satisfies for any $A, B \in \mathcal{P}(X)$,*

$$|A| = |B| \Rightarrow \mu(A) = \mu(B).$$

In the same spirit of k-additive measures, p-symmetric measures appear as a middle term between symmetric measures and general fuzzy measures. They reduce the complexity of fuzzy measures and provide a generalization of the idea of symmetry. The notion of p-symmetry is based in the concepts of indifferent elements and subsets of indifference.

Definition 9 *[40] Given two elements x_i, x_j of the universal set X, we say that x_i and x_j are* **indifferent elements** *for μ if and only if*

$$\forall A \subseteq X \backslash \{x_i, x_j\},\ \mu(A \cup x_i) = \mu(A \cup x_j).$$

In Multicriteria Decision Making, the definition of indifferent elements reflects the fact that criteria x_i and x_j are equivalent, so that we do not care about which one is fulfilled.

This can be extended to more than two elements through subsets of indifference.

Definition 10 *[40] Given a subset A of X, we say that A is a* **subset of indifference** *if and only if $\forall B_1, B_2 \subseteq A$, $|B_1| = |B_2|$ and $\forall C \subseteq X \backslash A$, it is*

$$\mu(B_1 \cup C) = \mu(B_2 \cup C).$$

From this definition, two elements in the same subset of indifference are indifferent elements in the sense of Definition 9.

Let Π_1, Π_2 be two partitions of X. We say that Π_2 is **coarser** than Π_1, denoted $\Pi_1 \sqsubseteq \Pi_2$ if

$$\forall A_1 \in \Pi_1,\ \exists A_2 \in \Pi_2,\ \text{ s.t. } A_1 \subseteq A_2.$$

Now, it is straightforward to define p-symmetric measures.

Definition 11 *[40] Given a fuzzy measure μ, we say that μ is a p-***symmetric measure** *if and only if the coarsest partition of the universal set in subsets of indifference is $\{A_1, ..., A_p\}, A_i \neq \emptyset, \forall i \in \{1, ..., p\}$.*

With these definitions, a symmetric measure is just a 1-symmetric measure. Given $\{A_1, ..., A_p\}$ a partition of X, the set of all fuzzy measures μ such that A_i for any $i = 1, ..., p$, is a subset of indifference for μ is denoted by $\mathcal{FM}(A_1, ..., A_p)$. Remark that $\mathcal{FM}(A_1, ..., A_p)$ is a convex polyhedron for fixed $A_1, ..., A_p$.

When dealing with a p-symmetric measure w.r.t. $\{A_1, ..., A_p\}$, we only need to know the number of elements of each A_i that belong to a given subset B of the universal set X. Then, we can identify $B \subseteq X$ with a p-dimensional vector $(b_1, ..., b_p)$ where $b_i := |A_i \cap B|$, $\forall i = 1, ..., p$.

This property allows a reduction in the complexity of the measure:

Lemma 1 *[41] Let μ be a p-symmetric measure w.r.t. the indifference partition $\{A_1, ..., A_p\}$. Then, it can be represented in a $(|A_1| + 1) \times \cdots \times (|A_p| + 1)$ matrix whose coefficients are defined by*

$$M(i_1, ..., i_p) := \mu(i_1, ..., i_p), i_j \in \{0, ..., |A_j|\}.$$

Remark 2 *Lemma 1 also holds for any measure in $\mathcal{FM}(A_1, ..., A_p)$.*

As pointed out by Valaskova [52], $A_1, \ldots, A_p$ are the equivalance classes of the indifference equivalence relation. In [41], some other interesting properties of p-symmetric measures have been studied.

Consider a given object f, whose corresponding scores on each criterium are $f(x_1), \ldots, f(x_n)$, which are numerical values. To compare different objects, we need to obtain an overal score from $f(x_1), \ldots, f(x_n)$. This is done through an aggregation operator [34], the Choquet integral [11] being among the most popular.

Definition 12 *The* **Choquet integral** *of a function*

$$f : X \to [0, 1]$$

with respect to a fuzzy measure μ on X is defined by

$$\mathcal{C}_\mu(f) := \sum_{i=1}^{n} (f(x_{(i)}) - f(x_{(i-1)}))\mu(B_i),$$

where $\{x_{(1)}, \ldots, x_{(n)}\}$ is a permutation of the set $\{x_1, \ldots, x_n\}$ satisfying

$$0 = f(x_{(0)}) \leq f(x_{(1)}) \leq ... \leq f(x_{(n)}),$$

and

$$B_i = \{x_{(i)}, ..., x_{(n)}\}.$$

The Choquet integral is a generalization of the concept of expected value, and models based on it are generalizations of the expected utility model.

With all these concepts and notations, our problem can be stated as follows: Consider m objects represented by the functions $f_1, \ldots, f_m$. Assume the aggregation operator applied to obtain an overall score is the Choquet integral and that the corresponding value of function f_i is y_i, $i = 1, \ldots, m$. Consider a subfamily of fuzzy measures $\mathcal{F}$. We look for $\mu \in \mathcal{F}$ minimizing

$$\sum_{i=1}^{m} (\mathcal{C}_\mu(f_i) - y_i)^2. \tag{2}$$

That is, we look for the fuzzy measure μ in $\mathcal{F}$ that best fits our data, with the squared error as criterion of fitness.

3 Genetic algorithms

The problem of identification of fuzzy measures from sample data (stated in previous section) can be written as a quadratic problem if $\mathcal{F}$ is convex (see [28]) and then solved with the usual methods. This approach will always lead to the exact solution, but it can be very time-consuming and has a strong tendency to overfitting, thus leading to bad approximations when there exists some noise in the data.

For these reasons, several alternative methods have been proposed. Usually, they are based in heuristics and lead to sub-optimal results, but are much more convenient in practice. For instance, we have the methods proposed by Grabisch [24], Mori and Murofushi [43] or Beliakov et al. [8]. Wang [55] has also proposed the use of genetic algorithm to solve the problem. We will follow this approach in this paper, but proposing a new cross-over operator (cf. Section 4). In this section we briefly describe genetic or evolutionary algorithms, in order to be self-contained.

Genetic algorithms are general optimization methods based on the theory of natural evolution [23, 32]. The main concepts are those of *individual* and *population*, which are, respectively, a candidate solution and the set of individuals being considered at a certain step in the algorithm.

Starting from an initial population, at each iteration (or *generation*), some individuals are selected with probability proportional to their *fitness* (which is measured according to the function that we want to optimize) and new individuals are generated from them using a *cross-over operator*. These new individuals replace the old ones (their *parents*) and the process continues till an optimum is found or till the maximum number of generations is reached (or other suitable termination condition holds). Then, the best individual in the last population is returned as a possible solution to the problem.

Sometimes, the cross-over operator greatly reduces the diversity of the populations and the risk of finding only a local optimum increases. To avoid this, a *mutation operator* which randomly changes individuals is defined. With some predetermined frequency, this operator is applied to individuals chosen at random from the population. The whole process is summarized in Algorithm 1.

Algorithm 1 A basic genetic algorithm

Generate initial population
repeat
  Evaluate fitness of every individual in the population
  Select individuals to reproduce
  Mate pairs of individuals and apply cross-over operator
  Select individuals to mutate and apply mutation operator
until termination condition is reached

There exist a number of variants to this basic genetic algorithm. For instance, one can maintain several populations in parallel and migrate individuals from one to another in each generation. This is usually known with the name of *deme genetic algorithm*. Also, one can choose among several possibilities the way the new population is generated. In a *simple genetic algorithm* the whole population is replaced by individuals generated from the old one by cross-over (one can choose to always keep the best individuals of the old population, and then the algorithm is said to be *elitist*); in *steady-state genetic algorithms* only a proportion of the old population (usually, the individuals with worst fitness) is replaced by the newly generated offspring

(in this case one says the populations are *overlapping*). *Incremental genetic algorithms* constitute an extreme case, in which only one or two new individuals are produced in each generation.

There exist several different ways of selecting the individuals which will be mated for cross-over. Some of the most used are:

- *Roulette wheel selection or stochastic sampling with replacement*: An individual has probability p of being selected, where p is equal to the proportion of its fitness to the sum of all the fitness of the individuals of the population.
- *Stochastic sampling without replacement*: Similar to roulette wheel selection, after each selection the individual is removed from the original population and the probabilities are recalculated.
- *Tournament selection*: Two individuals are selected according to the roulette wheel selection and the one with highest fitness is chosen.
- *Deterministic sampling*: Each individual is selected for cross-over a number of times equal to the integer part of $p \times n$, where n is the size of the population and p is as in roulette wheel selection.
- *Uniform selection*: Every individual has the same probability of being selected.

In some cases a scaling is applied to the fitness of the individuals before the selection is performed, in order to avoid preponderance of certain individuals with very high scores (possibly local optima). This scaling can be linear, follow a power law, truncate the values at certain thresholds, ... (for more details see [23]).

These algorithms have been successfully used in many optimization problems (see [23]) and have a lot of good properties (robustness, few requirements on the function to optimize, low complexity, ...) which make them an appealing choice for the problem of identification.

4 The algorithm

Suppose we have functions $f_1, \ldots, f_m$ on n criteria and $y_1, \ldots, y_m$ are their corresponding Choquet values (probably affected by some noise). We want to find a fuzzy measure $\mu \in \mathcal{F}$ minimizing the quadratic error (expression 2).

To apply the method of genetic algorithms to our problem, we must first choose a cross-over operator on fuzzy measures and select a suitable representation. In previous works on this topic (see, for instance, [55]), the cross-over operator might lead to measures which are not monotone, so this condition must be checked every time that the operator is applied and, if it does not hold, the coefficients of the measure have to be modified in order to ensure monotonicity. This decreases the efficiency of the algorithm.

To overcome this drawback, we propose to use as cross-over operator the *convex combination* of two fuzzy measures μ_1 and μ_2, defined by

$$\lambda\mu_1 + (1-\lambda)\mu_2,$$

with $\lambda \in [0,1]$ chosen at random when the operator is applied.

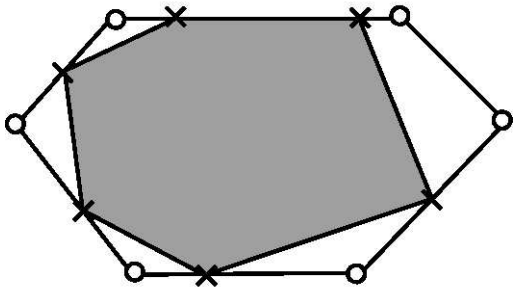
It is clear that the convex combination of any two fuzzy measures is also a fuzzy measure, so we do not check monotonicity of the resulting measure after each application of the cross-over operator. Another advantage of this operator comes from the fact that it can also be applied to subfamilies of fuzzy measures which are convex, such as k-additive measures and p-symmetric measures (when the partition of subsets of indifference is fixed). Then, we can select different classes of measures for the learning problem without having to define new, specific cross-over operators.

The main problem with this operator is that the search space is reduced in each generation (see Figure 1), since the convex combinations of the new measures are always a subset of the convex combinations of the previous ones. What is more, if the initial population is not carefully selected, then it might be the case that only bad approximations could be found among the convex combinations of those initial individuals.

To deal with this problem, one can use the extreme points of the subfamily (see Section 5). Then, the starting population is initialized to the set of vertices and, seldom, some individuals are combined with one of the vertices chosen at random (this is the mutation operator) in order to keep the diversity high enough.

Once we know the vertices, the natural representation of the measures is by their coefficients as convex combination of the initial population. With this representation the cross-over operator has low complexity (if the number of vertices is small) and also the number of Choquet integrals that must be computed is kept to a minimum, since they are also convex combinations of the integrals of the parents. Thus, we only have to integrate with respect to the vertices, which is usually easy to compute. Although the GA could be time-comsuming if the termination condition is not reached in after a moderated number of iterations, these

Figure 1: Reduction of search space due to convex combination



optimizations in the representation of the measures (together with imposing a maximum number of iterations as stopping criterion) makes the mean computing complexity be not high.

Then, a measure μ will be represented as

$$(\lambda_1, \ldots, \lambda_l, z_1, \ldots, z_m)$$

where $\lambda_1, \ldots, \lambda_l$ are the coefficients of μ with respect to the l vertices and $z_i = \mathcal{C}_\mu(f_i)$. From this representation the fitness function (quadratic error) can be easily computed too, since it is simply

$$\sum_{i=1}^{m} (z_i - y_i)^2.$$

We have implemented this algorithm in a program[1] called *fmlearner* (for fuzzy measure learner). When using *fmlearner*, the user can select among different subfamilies of fuzzy measures for the learning (general fuzzy measures, k-additive measures and p-symmetric measures). It is also possible to select a wide number of parameters which affect the evolution of the populations, like the kind of genetic algorithm used, the number of generations, the probability of cross-over and mutation, the way of selecting individuals for reproduction and the scaling of fitness values (cf. Section 3).

5 Extremes of the fuzzy measures

As has been noted in the previous section, when using the convex combination as cross-over operator in our algorithm, the only way to guarantee that all possible measures are inside the search region consists in setting the initial population to the set of extreme points of the fuzzy measures.

Determining these extremes turns out to be an interesting theoretical problem on its own. He have addressed this question in [39], where we have proved the following results.

Proposition 1 *The set of $\{0,1\}$-valued measures constitutes the set of vertices of $\mathcal{FM}$.*

Theorem 1 *The set of extreme points of $\mathcal{FM}(A_1, ..., A_p)$ is the set of $\{0,1\}$-valued measures that are also in $\mathcal{FM}(A_1, ..., A_p)$.*

The case of k-additive measures is much more difficult to handle. For probabilities, i.e. 1-additive measures, the following result holds:

Proposition 2 *The extreme points of $\mathcal{FM}^1$ are the $\{0,1\}$-valued measures that are in $\mathcal{FM}^1$. These $\{0,1\}$-valued measures are the unanimity games on the singletons u_{x_i}, $x_i \in X$. Moreover, given P a probability distribution over X, it can be written as*

$$P = \sum_{x_i \in X} P(x_i) u_{x_i}.$$

Let us now turn to the special case of 2-additive measures. The following can be proved:

[1] We have used the GAlib genetic algorithm package, written by Matthew Wall at the Massachusetts Institute of Technology.

Proposition 3 *The set of extreme points of $\mathcal{FM}^2$ are the $\{0,1\}$-valued measures that are in $\mathcal{FM}^2$. These $\{0,1\}$-valued measures are given by:*

- $m(x_i)=1$, $m(A)=0$, *otherwise (the extreme points of probabilities, u_{x_i}, $x_i \in X$).*
- $m(x_i)=1$, $m(x_j)=1$, $m(x_i,x_j)=-1$, $m(A)=0$, *otherwise. We will denote these measures by μ'_{x_i,x_j}.*
- $m(x_i,x_j)=1$, $m(A)=0$, *otherwise (u_{x_i,x_j}, $\{x_i,x_j\} \subseteq X$).*

Moreover, given μ a 2-additive measure over X, it can be written as

$$\mu = \sum_{m(x_i,x_j)<0} -m(x_i,x_j)\mu'_{x_i,x_j} + \sum_{m(x_i,x_j)>0} m(x_i,x_j)u_{x_i,x_j} + \sum_{x_i \in X} c(x_i)u_{x_i},$$

with $c(x_i) := [m(x_i) + \sum_{m(x_i,x_j)<0} m(x_i,x_j)]$.

Remark 3 *Notice that $u_{x_i,x_j} = u_{x_i}u_{x_j}$ and that $\mu'_{x_i,x_j} = u_{x_i} \vee u_{x_j}$.*

However, these results cannot be extended when $k > 2$, as the next theorem shows.

Theorem 2 *There are vertices of the set $\mathcal{FM}^k, k > 2$, that are not $\{0,1\}$-valued measures.*

The determination of the vertices of k-additive measures with $k > 2$ is an interesting open problem, and so it is the calculation of their number, which we will consider in Section 7.

6 The experiments

To test the performance of our algorithm in the problem of identification we adopt the procedure followed by Grabisch in [24]. Namely, we consider the fuzzy measure whose values are presented in Table 1. The input to the algorithm are the values of the measure on the 81 points of the form (x_1,x_2,x_3,x_4) with $x_i \in \{0, 0.5, 1\}$ for $i = 1,\ldots,4$.

A	$\mu(A)$	A	$\mu(A)$	A	$\mu(A)$
$\{1\}$	0.1	$\{1,2\}$	0.3	$\{1,2,3\}$	0.5
$\{2\}$	0.2105	$\{1,3\}$	0.3235	$\{1,2,4\}$	0.8667
$\{3\}$	0.2353	$\{1,4\}$	0.7333	$\{1,3,4\}$	0.8824
$\{4\}$	0.6667	$\{2,3\}$	0.4211	$\{2,3,4\}$	0.9474
		$\{2,4\}$	0.8070		
		$\{3,4\}$	0.8235		

Table 1: Values of the measure

To simulate the situation in practice, where the sample data are often affected by some noise, we add to these values a Gaussian noise of increasing variance ($\sigma^2 = 0.0, 0.00096, 0.00125, 0.00625, 0.0125$), thus obtaining 5 different identification problems (of the same original measure).

This is repeated 100 times to reduce the influence of the random selections appearing in the application of the method and the effect of random modification of examples by the noise.

We have performed the experiments with the 5 different families of measures that have been included in our implementation of the algorithm. These are general fuzzy measures, k-additive measures (in the experiments we chose $k = 2$), p-symmetric measures (again we chose $p = 2$ for these experiments), probabilities and symmetric measures.

As explained before, our implementation allows the user to select a wide number of parameters which affect the evolution of the populations. The values of the parameters selected for these experiments are presented in Table 2.

To evaluate the error of the approximation we compute the average quadratic error on the considered data sample (that is, the original data, not affected by the noise). Then, if our algorithm gives us the measure μ' as approximation of the original measure μ, the error is defined by

$$\frac{1}{m}\sum_{i=1}^{m}(\mathcal{C}_{\mu'}(f_i) - \mathcal{C}_{\mu}(f_i))^2, \tag{3}$$

Parameter	Value
Number of generations	1000
Number of parallel populations	10
Probability of cross-over	0.99
Probability of mutation	0.01
Type of selector	Roulette wheel
Scaling of fitness scores	No

Table 2: Values of the parameters

where $\{f_i\}_{i=1,\ldots,m}$ are the m objects considered as examples (in our case, $m = 81$).

With this expression, we can directly compare the error really obtained (Equation 3) with the error expected if the approximation of the algorithm is perfect, that is $\mathcal{C}_{\mu'}(f_i) = y_i$, where y_i is the original value $\mathcal{C}_\mu(f_i)$ plus the Gaussian noise N_i. In that case the error would be

$$\frac{1}{m}\sum_{i=1}^{m}(\mathcal{C}_{\mu'}(f_i) - \mathcal{C}_\mu(f_i))^2 = \frac{1}{m}\sum_{i=1}^{m}(y_i - \mathcal{C}_\mu(f_i))^2 = \frac{1}{m}\sum_{i=1}^{m}N_i^2,$$

which in average is σ^2, the variance of the noise.

The average error of the 100 executions for the different subfamilies used in the approximation and the different values of σ is presented in table 3.

σ^2\family	2-additive	2-symmetric	Probability	Symmetric	General
0.0	5.13779E-05	0.0173771	0.00248447	0.0256926	0.00141472
0.00096	0.00012567	0.0172276	0.00252335	0.0257424	0.0014723
0.00125	0.000152595	0.0169306	0.00252909	0.025766	0.00141241
0.00625	0.000511731	0.0176761	0.00270966	0.0259294	0.00183267
0.01250	0.00100916	0.0182342	0.00300766	0.0261342	0.00241865

Table 3: Average error of the approximation

We observe that the best approximation in this experiment is obtained when using 2-additive measures. This may be caused by the fact that the measure to be approximated is close to be 2-additive (the biggest value of its Möbius transform on sets with more than two elements is reached at $m(\{2,3,4\}) = 0.0083$). With this subfamily the error of the approximation increases with the variance of the noise added to the data (as expected). However, for non-null noise, it is always less than σ^2. Therefore, it seems that the deviations in the data are not amplified by our algorithm (but rather decreased).

In Table 4 we compare the results obtained with our algorithm when using 2-additive measures, with those obtained with the algorithms of Grabisch [24] and of Mori and Murofushi [43]. The results of these algorithms when applied to this problem are taken from [24].

σ^2\Algorithm	Mori y Murofushi	Grabisch	fmlearner
0.0	0.0000	1.4E-7	5.13779E-05
0.00096	0.00087	0.00083	0.00012567
0.00125	0.0117	0.0108	0.000152595
0.00625	0.0605	0.0530	0.000511731
0.01250	0.1211	0.1054	0.00100916

Table 4: Comparison with other algorithms

From the table it can be observed that, although for exact sample data the other algorithms could be preferable, our algorithm seems to perform consistently better when noise is present. This is a desirable and convenient property, since in practical situations it is impossible to avoid the existence of noise, specially if only ordinal information is available.

7 Alternative initializations

Despite the fact that the only way to guarantee that the solution of the learning problem lays into the search region is to set the initial population of the algorithm to be the set of vertices, there is a tremendous drawback with this approach: The number of vertices of general fuzzy measures on n criteria coincides with the number of non-empty antichains (collections of sets which are pairwise uncomparable with respect to inclusion, see [1]) on a set of n elements. This is easily seen, since every vertex is a $\{0,1\}$-valued measure (Proposition 1) and vice versa, and every such measure is determined by its minimal sets (and vice versa), which, obviously, form an antichain.

The sequence formed by the number of different antichains or *Sperner systems* on a set of n elements is known with the name of the sequence of *Dedekind numbers* (see [15]). The form of the general term of this sequence is not known, and in fact, the only Dedekind numbers which have been calculated up today are the first 8 (the eighth Dedekind number was first computed by Wiedemann [56]). They are presented in Table 5.

n	Dedekind numbers
1	1
2	4
3	18
4	166
5	7579
6	7828352
7	2414682040996
8	56130437228687557907786

Table 5: Number of vertices

Though the whole sequence is not known, its asymptotical behavior is, as the following result shows.

Theorem 3 *[30, 21] If D_n is the n-th Dedekind number, then it holds*

$$2^q \leq D_n \leq 3^q$$

with $q = \binom{n}{\lfloor \frac{n}{2} \rfloor}$ and $\lfloor x \rfloor$ the integer part of x, for all $n \geq 1$.

There exist closer bounds (see [13]), but these ones, together with Table 5 show that the use of all vertices of the fuzzy measures as initial population in our algorithm is infeasible in practice.

Thus, it is needed to develop alternative and feasible methods of initialization. We propose two different methods: the mincut method and a random initialization.

7.1 Limitation of the number of minimal sets (mincuts)

A natural way of classifying the vertices of the fuzzy measures is by the number of minimal subsets. In fact, these vertices can be seen as monotone boolean functions and this kind of classification is usual in their study [35, 49] and has significant practical relevance [6, 31]. In that context the minimal subsets of a function are known with name of *mincuts* or *lower units*, and they correspond to the terms which form the expression of the function in their minimal Disjunctive Normal Form [22].

Then, we propose to study the influence in the performance of the method of limiting the number of mincuts in the vertices of the initial population. This will reduce the size of the vertices considered, as can be seen from the following result.

Theorem 4 *[45] If $\alpha(n,m)$ is the number of monotone boolean functions on n variables with exactly m mincuts, then it holds*

$$\begin{aligned}
&\alpha(n,1) = 2^n \\
&\alpha(n,2) = \tfrac{1}{2}(2^n)(2^n-1) - (3^n - 2^n) \\
&\alpha(n,3) = \tfrac{1}{6}(2^n)(2^n-1)(2^n-2) - (6^n - 5^n - 4^n + 3^n)
\end{aligned}$$

for $n > 1$.

The values of $\alpha(n,m)$ for $3 < m \leq 10$ are also known due to Cvetković [14] ($m = 4$), Arocha [4, 5] ($m = 5, 6$) and Kilibarda and Jovović [33], which gave a general procedure for calculating these numbers for n arbitrary and $1 \leq m \leq 10$. However, we are only interested in noticing that the number of functions with a fixed quantity of mincuts grows much slowly than the number of all vertices (though still exponentially).

Thus, we propose to use an alternative to the set of all vertices in the initialization of the algorithm: the set of all vertices with up to m mincuts. The maximum value of m for a given value of n is given by Sperner's lemma:

Theorem 5 Sperner's lemma *[1] If $\mathcal{S}$ is an antichain of a set with n elements then*

$$|\mathcal{S}| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$$

for $n \geq 1$, where $|\mathcal{S}|$ denotes the number of subsets in $\mathcal{S}$.

To test this procedure, we have conducted the experiments described in Section 6, but now initializing the algorithm with the vertices of fuzzy measures which have up to m mincuts, with $m = 1, .., 6$ (notice that $6 = \binom{4}{2}$, the maximum size of an antichain of a set of 4 elements). The number of vertices used in each case is shown in Table 6. The rest of parameters are, again, those described in Table 2. The result of these experiments (average quadratic error of the 100 executions) is presented in Table 7.

Mincuts	Number of vertices
1	15
2	70
3	134
4	159
5	165
6	166

Table 6: Number of vertices on 4 criteria with up to m mincuts

σ^2\Mincuts	1	2	3	4	5	6
0.0	0.00340392	0.000912161	0.00110968	0.00130424	0.00137734	0.00138689
0.00096	0.00340387	0.000954684	0.00118035	0.00140243	0.00139888	0.00146325
0.00125	0.0035783	0.000993236	0.00119306	0.00137859	0.00144252	0.00144927
0.00625	0.00361389	0.00140584	0.0016565	0.0018119	0.00191207	0.00193089
0.01250	0.0040022	0.00193432	0.00204001	0.00230163	0.00230736	0.00238539

Table 7: Average error of the approximation

It can be observed that, though using less than half of the total number of vertices, the approximations obtained when restricting the mincuts of the extreme points to 2 are the best ones in these experiments. Also with mincuts ranging from 3 to 5 we obtain better performance than when using all the vertices (6 mincuts). However, restricting to 1 the number of mincuts yields poorer results.

7.2 Random initialization

The experiments in the previous section suggest that, in some situations, the restriction of the number of mincuts of the vertices does not worsen the performance of the algorithm (and it can even improve it), while reduces the number of extreme points that ought to be considered. However, notice that for a number of criteria not small, the number of vertices can grow very large even with the proposed restrictions. For instance, if we have n criteria, the number of vertices with exactly one mincut are, obviously, $2^n - 1$. Consequently, we need to develop another initialization of the population that allows to keep its size reduced even when the number of criteria is large.

To this extent, we study an approach which is usual when applying genetic algorithms to solve a problem: a random initialization of the population. With this kind of initialization we can always choose the number of measures that will form the initial population and this number will not depend on the number of criteria.

Although in this case it is impossible to guarantee that a solution to the problem can be found inside the search space (the same happened when restricting the number of minimal subsets of the vertices), we want to explore whether this method is useful in practical situations. Also, we are interested in studying in which situations it is mandatory to use the extreme points in order to obtain an acceptable approximation.

The method that we propose to obtain random fuzzy measures is presented in Algorithm 2.

Algorithm 2 Random generation of a fuzzy measure μ on a set X

$\mu(\emptyset) := 0$
$\mu(X) := 1$
while there exists $A \subset X$ such that $\mu(A)$ is undefined **do**
  randomly choose $A \subset X$ such that $\mu(A)$ is undefined
  $min := 0$
  $max := 1$
  for every $B \subset A$ such that $\mu(B)$ is defined **do**
    if $\mu(B) > min$ **then**
      $min := \mu(B)$
    end if
  end for
  for every $B \supset A$ such that $\mu(B)$ is defined **do**
    if $\mu(B) < max$ **then**
      $max := \mu(B)$
    end if
  end for
  $\mu(A) :=$ random value between min and max
end while

This random generation has been tested experimentally and we have found that if a sequence of measures μ_i, $i = 1, \ldots, p$ is generated in this way, then the average measure

$$\frac{1}{p} \sum_{i=1}^{p} \mu_i$$

is approximately the measure whose Choquet integral is the arithmetic mean of the values on the criteria (when p is large). This seems a natural and desirable property for a random generator of fuzzy measures.

We have repeated the experiments for the learning of the measure in Table 1 with the settings described in Section 6. Again, the parameters of the algorithm are those presented in Table 2. The number of measures considered in the initial population is 100, which is less than the total number of vertices (166 in this case).

The average quadratic error of the 100 executions is shown in Table 8. These results are comparable to those obtained when using all the vertices or a restricted number of them (see Table 7).

σ^2	Random initialization
0.0	0.00115884
0.00096	0.00136715
0.00125	0.00129024
0.00625	0.00156613
0.01250	0.00199907

Table 8: Average error of the approximation

The method of random initialization presented in this section is only applicable to generate general fuzzy measures. When using other subfamilies, other generation algorithms should be used. As far as we know, in the literature this problem has not been addressed yet.

7.3 Random initialization versus vertices

In the previous section we have observed that the performance of the algorithm when using either random initialization or vertices is similar in the experiment proposed in [24]. However, it seems logical that each kind of initialization will perform better in different regions of the search space. When the solution to the problem is close to one extreme point, the vertices should perform better, while the random generation might offer better results when the measure to be approximated is far away from every vertex (since less generations would be needed to get close to the solution).

To test this hypothesis we have conducted experiments with exactly the same settings as in previous sections, but now trying to approximate measures different from that described in Table 1. The first measure is the one whose Choquet integral is the arithmetic mean of the criteria values (note that in Section 7.2 we have noticed that this measure is the average of the random measures generated with Algorithm 2). The average quadratic errors of the results of the experiment are presented in Table 9.

σ^2\Initialization	Random	1 mincut	2 mincuts	3 mincuts	4 mincuts	5 mincuts	6 mincuts
0.0	3.94165e-05	0.00108329	0.00105597	0.0012329	0.00131521	0.00148376	0.00149476
0.00096	0.000146103	0.00111882	0.00114657	0.00138279	0.00152615	0.00165256	0.00155148
0.00125	0.000167288	0.00109699	0.0011401	0.00137238	0.00160874	0.00166164	0.00163334
0.00625	0.000637256	0.00135049	0.0016517	0.0018606	0.00199438	0.00198621	0.00207653
0.01250	0.0010957	0.00178214	0.00216861	0.00228676	0.00244426	0.00253897	0.00251612

Table 9: Average error of the approximation

It is clear that in this situation the random initialization seems to offer better results, as expected.

On the other hand, we have performed experiments trying to approximate measures which are close to the vertices. We also want to test whether all the vertices are necessary, so we have chosen measures which are close to vertices with mincuts ranging from 1 to 6. The way of constructing the measures to be approximated is simple. We take one vertex μ and consider the measure μ' defined by:

$$\mu'(A) = \begin{cases} 0 & \text{if } \mu(A) = 0 \\ 0.95 & \text{if } \mu(A) = 1 \text{ and } A \neq X \\ 1 & \text{if } A = X \end{cases}$$

The vertices selected for generating the measures μ' are those whose minimal subsets are presented in Table 10.

Vertex	Minimal sets
1	$\{1\}$
2	$\{1\}, \{2,3\}$
3	$\{1,2,3\}, \{1,2,4\}, \{2,3,4\}$
4	$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3,4\}$
5	$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}$
6	$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}$

Table 10: Minimal sets of the vertices considered

The average quadratic errors of approximation obtained in the experiments are shown in Table 11. For lack of space, we present only the results achieved when no noise was added to original data. When noise was considered, the trends remained the same.

As it could be expected, in all these cases the random initialization performs very poorly, while the extremes get the better results. Even more, one can observe that for the approximation of the functions obtained from vertices with m mincuts it is not enough to use extreme points with a smaller number of minimal sets. This suggests that all the vertices are needed in certain situations.

8 Conclusions and future work

We have introduced a method for the identification of fuzzy measures from sample data using genetic algorithms. The convex combination was selected as cross-over operator so that the resulting algorithm can be

Vertex\Init.	Random	1 mincut	2 mincuts	3 mincuts	4 mincuts	5 mincuts	6 mincuts
1	0.0225958	1.17156e-05	0.000107482	0.000160329	0.00017053	0.000178684	0.000177436
2	0.0201573	0.0372242	0.000127292	0.000187425	0.000219845	0.000208076	0.000209426
3	0.0196933	0.0159139	0.004174	3.41339e-05	3.66318e-05	3.94463e-05	4.03799e-05
4	0.0197049	0.0311471	0.0123082	0.00484678	0.000146568	0.000151382	0.000140635
5	0.0194112	0.0431124	0.0166618	0.00803647	0.00371792	0.00020667	0.000205906
6	0.0167142	0.0481669	0.0205665	0.0113583	0.0056384	0.00290944	0.000227081

Table 11: Average error of the approximation

used with different subfamilies of measures. The results of the experiments that we have carried out suggest that the method is much more stable with respect to the presence of noise in the sample data than other existing algorithms.

However, in order to ensure a good performance of our algorithm, we need to consider as initial population the set of all vertices of the considered subfamily. We have shown that the number of vertices grows very quickly with the cardinality of the universal set. To deal with this problem, we have also proposed two different alternatives: a restriction in the number of vertices considered an a method for generating random measures. Both seem to work well in normal situations, but under some circumstances (when the measure to approximate is near the vertices or far away from them) one can perform better than the other.

There is a number of open problems that we would like to study in the near future. First of all, we would like to investigate deeply the influence of the different parameters (number of examples, family of measures,...) on the performance of the method.

Following the ideas presented in this paper, it would be interesting to extend the methods of random initialization and of reduction of vertices to other subfamilies of measures such as k-additive measures or p-symmetric measures. If the method of mincuts of Section 7.1 is used, another interesting open problem is to determine the best choice of m for a given set of data. We also want to study further the properties of the random generator of measures that we have introduced. In this sense, when dealing with a given subfamily, we have to search for a random generation method leading to a "uniform" selection in the subfamily.

Also, it would be interesting to study the performance of the method when other subfamilies of fuzzy measures are used for the approximation. A promising case is that of the k-additive beliefs [17], that are very interesting in Evidence Theory [48], and whose extreme points correspond to the Dirac probabilities.

Finally, the results of the experiments presented in Section 7.3 show that it is necessary to study the conditions under which the set of vertices (with a certain number of mincuts) is needed for the initialization of the algorithm. It seems that an adequate notion of "distance to the vertices" may be a good start-point.

Acknowledgements

This research has been supported in part by FEDER-MCYT grant number BFM2001-3515.

References

[1] M. Aigner. *Combinatorial Theory*. Springer-Verlag, 1979.

[2] M. Allais. Le comportement de l'homme rationnel devant le risque: critique des postulats de l'école américaine. *Econometrica*, (21):503–546, 1953. In French.

[3] F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, (34):199–205, 1963.

[4] J. L. Arocha. The number of antichains of the length 5 and 6. Technical report, VINITI, 1982. In Russian.

[5] J. L. Arocha. Anticadenas en conjuntos ordenados (Antichains on ordered sets). *An. Inst. Mat. Univ. Nac. Autónoma México*, 27:1–21, 1987. In Spanish.

[6] J. Astola and P. Kuosmanen. *Fundamentals of Nonlinear Digital Filtering*. CRC Press, 1997.

[7] M. Bazaraa, J. Jarvis, and H. Sherali. *Linear programming and network flows*. John Wiley and Sons, 1990.

[8] G. Beliakov, R. Mesiar, and L. Valaskova. Fitting generated aggregation operators to empirical data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(2):219–236, 2004.

[9] A. Chateauneuf. Modelling attitudes towards uncertainty and risk through the use of Choquet integral. *Annals of Operations Research*, (52):3–20, 1994.

[10] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, (17):263–283, 1989.

[11] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, (5):131–295, 1953.

[12] E. F. Combarro and P. Miranda. A genetic algorithm for the identification of fuzzy measures from sample data. In *Proceedings of International Fuzzy Systems Association Conference (IFSA'03)*, pages 163–166, Istambul (Turkey), 2003.

[13] J. Crampton. *Authorization and antichains*. PhD thesis, Birkbeck, University of London, London, England, 2002.

[14] D. Cvetković. The number of antichains of finite power sets. *Publ. Inst. Math. (Beograd)*, 13(27):5–9, 1972.

[15] R. Dedekind. Über Zerlegungen von Zahlen durch ihre grössten gemeinsamen Teiler. *Festschrift Hoch Braunschweig Ges. Werke*, II:103–148, 1897. In German.

[16] D. Denneberg. *Non-additive measures and integral*. Kluwer Academic, 1994.

[17] D. Denneberg. Non-additive measure and integral, basic concepts and their role for applications. In T. Murofushi M. Grabisch and M. Sugeno, editors, *Fuzzy measures and integrals. Theory and applications*, volume 40 of *Studies in Fuzziness and Soft Computing*, pages 42–69. Physica-Verlag, 2000.

[18] D. Dubois, H. Prade, and R. Sabbadin. Qualitative decision theory with Sugeno integrals. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy measures and integrals*, number 40 in Studies in Fuzziness and Soft Computing, pages 314–331. Physica-Verlag, 2000.

[19] C. A. Bana e Costa and J.-C. Vansnick. MACBETH- An interative path towards the construction of cardinal value function. *International Transactions in Operational Research*, 1(4):489–500, 1994.

[20] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.*, (75):643–669, 1961.

[21] K. Engel. *Sperner Theory*. Cambridge U. Press, 1997.

[22] E. Gilbert. Lattice-theoretic properties of frontal switching functions. *J. Math. Phys.*, 33:57–67, 1954.

[23] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, 1989.

[24] M. Grabisch. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Proceedings of Int. Joint Conf. of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Sysmposium*, pages 145–150, Yokohama (Japan), March 1995.

[25] M. Grabisch. k-order additive discrete fuzzy measures. In *Proceedings of 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1345–1350, Granada (Spain), 1996.

[26] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:587–607, 1997.

[27] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, (92):167–189, 1997.

[28] M. Grabisch and J.-M. Nicolas. Classification by fuzzy integral-performance and tests. *Fuzzy Sets and Systems, Special Issue on Pattern Recognition*, (65):255–271, 1994.

[29] P. L. Hammer and R. Holzman. On approximations of pseudo-boolean functions. *Zeitschrift für Operations Research. Mathematical Methods of Operations Research*, (36):3–21, 1992.

[30] G. Hansel. Sur le nombre des fonctions booléennes monotones de n variables. *C. R. de l'Académie des Sciences de Paris*, 262(20):1088–1090, 1966.

[31] K. Hayase and H. Imai. OBDDs of a monotone function and of its prime implicants. In *Algorithms and Computation*, volume 1178 of *Lecture Notes in Computer Science*, pages 136–145. Springer, 1996.

[32] J. H. Holland. *Adaptation in natural and artificial systems*. Ann Arbor: The University of Michigan Press, 1975.

[33] G. Kilibarda and V. Jovović. On the number of monotone boolean functions with fixed number of lower units. *Intellektualnye sistemy*, 7:193–217, 2003. In Russian.

[34] G. Klir and T. A. Folger. *Fuzzy Sets, uncertainty and information*. Prentice Hall, 1989.

[35] A. D. Korshunov and I. Shmulevich. On the distribution of the number of monotone boolean functions relative to the number of lower units. *Discrete Mathematics*, 257:463–479, 2002.

[36] J.-L. Marichal. Tolerant or intolerant character of interacting criteria in aggregation by the Choquet integral. *European Journal of operational Research*, 155(3):771–791, June 2004.

[37] J.-L. Marichal, P. Meyer, and M. Roubens. Sorting multiattribute alternatives : the TOMASO method. *Computer and Operations Resarch*, to appear.

[38] R. Mesiar. Generalizations of k-order additive discrete fuzzy measures. *Fuzzy Sets and Systems*, 102(3):423–428, 1999.

[39] P. Miranda, E.F. Combarro, and P. Gil. Extreme points of some families of non-additive measures. Submitted.

[40] P. Miranda and M. Grabisch. p-symmetric fuzzy measures. In *Proceedings of Ninth International Conference of Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, pages 545–552, Annecy (France), July 2002.

[41] P. Miranda, M. Grabisch, and P. Gil. p-symmetric fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (Suppl.):105–123, 2002.

[42] P. Miranda, M. Grabisch, and P. Gil. Identification of non-additive measures from sample data. In R. Kruse G. Della Riccia, D. Dubois and H.-J. Lenz, editors, *Planning based on Decision Theory*, volume 472 of *CISM Courses and Lectures*, pages 43–63. Springer Verlag, 2003.

[43] T. Mori and T. Murofushi. An analysis of evaluation model using fuzzy measure and the Choquet integral. In *5th Fuzzy System Symposium*, pages 207–212, Kobe, Japan, 1989. In Japanese.

[44] G. Owen. *Game theory*. Academic Press, 1995.

[45] N. M. Riviere. Recursive formulas on free distributive lattices. *J. Combin. Theory*, 5:229–234, 1968.

[46] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (2):340–368, 1964.

[47] D. Schmeidler. Integral representation without additivity. *Proceedings of the American Mathematical Society*, (97(2)):255–261, 1986.

[48] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, (USA), 1976.

[49] Shapiro. On the counting problem for monotone boolean functions. *Comm. Pure Appl. Math.*, 23:299–312, 1970.

[50] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the theory of Games*, volume II of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, 1953.

[51] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[52] L. Valaskova. A note to the definition of p-symmetry. Working paper.

[53] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, New Jersey (USA), 1944.

[54] Z. Wang and G. Klir. *Fuzzy measure theory*. Plenum, New York, 1992.

[55] Z. Wang, K. Xu, J. Wang, and G. Klir. Using genetic algorithms to determine nonnegative monotone set functions for information fusion in environments with random perturbation. *International Journal of Intelligent Systems*, 14:949–962, 1999.

[56] D. Wiedemann. A computation of the eighth Dedekind number. *Order*, 8:5–6, 1991.