Tatra Mt. Math. Publ. **21** (2001), 133–152

TATRA MOUNTAINS
Mathematical Publications

A MEASURE OF THE DIFFERENCE BETWEEN TWO FUZZY PARTITIONS

SUSANA MONTES — PEDRO MIRANDA
— JORGE JIMÉNEZ — PEDRO GIL

ABSTRACT. The aim of this paper is to study the divergence measures between fuzzy partitions, which deal with measuring the difference of two partitions of a given finite universe. The interest of these measures is concentrated on the local case, in which the difference only depends on the difference of particular fuzzy subsets and on the measure of these subsets. This paper also states a necessary and sufficient condition for a divergence measure to be local. The whole study is based on previous results on fuzzy divergence measures between fuzzy subsets. In fact, the concept of divergence measure between partitions allows to obtain divergence measures between subsets and probabilistic divergence measures, as it is showed in the final sections. This paper is concluded by showing some important examples of families of divergence measures between fuzzy partitions.

1. Introduction

The purpose of our research consists on building up a framework for the most general kind of detection processes such as classify a population, compare some actual characteristics with the ideal ones, ..., which is a typical tool of decision support systems. We have considered the questionnaires ([10]) to do this. When classical detection is the object of the study, the goal, questions and answers are represented in crisp environment (crisp subsets and partitions). However, the crisp environment seems to be quite inappropriate to formulate questions and obtain answers in most of the cases of human or social interest. For example, the medical diagnosis is often based on questions as "have you got a bad headache?", which is a classical fuzzy proposition. In these cases, fuzzy theory and fuzzy logic

2000 Mathematics Subject Classification: 94A50, 94D05.
Keywords: divergence measures between fuzzy subsets, local divergence measures, fuzzy measures, divergence measures between partitions, probabilistic divergence measures.

The research in this paper has been supported in part by the DGES Grant No. PB97-1286 and a Research Grant of the *Fundación Banco Herrero*; its financial support is gratefully acknowledged. The authors also wish to express their heartily thanks to the referee for his careful reading and suggestions.

are more appropriate to describe the detection process. Fuzzy subsets and fuzzy partitions can better describe answers and questions as well as the goals of the questionnaires. So, in order to compare the actual state of the process with the desired classification classes, it seems to be quite natural to have a tool which measures to what extent the answers obtained so far fit one or more of the classes of interest. Since both answers and classes are (represented by) fuzzy subsets, and questions and goal are (represented by) fuzzy partitions, the ideal tool to make the comparisons seems to be the divergence measure between fuzzy sets and between fuzzy partitions. The smaller the divergence, the more similar the compared sets or partitions are, and zero divergence means perfect identification.

The problem we address is the following: we want to classify an object comparing it with some models. In order to do this, we study some characteristics (questions) of the object. The answers to the questions are fuzzy subsets. Now, we want to compare these fuzzy subsets with the models that are also fuzzy subsets; this comparison can be made by divergence measures. The models we are considering determine a fuzzy partition of the universal set. Of course, we can consider different fuzzy partitions. Which is the best one? To solve this problem we have to compare fuzzy partitions. In this paper we define divergence measures between fuzzy partitions as a tool to compare fuzzy partitions. Divergence measures between fuzzy partitions will be related to divergence measures between fuzzy subsets, as we will see.

2. Basic concepts

In this paper we are going to work with a classical measurable space formed by a referential Ω and a σ-field $\mathcal{A}$ formed by crisp subsets of Ω (in particular $\mathcal{A}$ can be the power crisp set of Ω). When Ω is a finite set, we will denote the elements of Ω by $x_1, \dots, x_n$. Fuzzy subsets are denoted by $A, B, \dots$ and also by $A_1, A_2, \dots$.

By using this σ-field $\mathcal{A}$ we define the set $\mathcal{A}^*$ formed by any fuzzy subset of Ω such that its membership function is measurable, when we work with the Borel σ-field. It can be proved that $\mathcal{A}^*$ is a σ-field over Ω formed by fuzzy subsets and also that the intersection between $\mathcal{A}^*$ and the power set of Ω is equal to the σ-field $\mathcal{A}$, that is, any elements of $\mathcal{A}$ belong to $\mathcal{A}^*$ and any crisp elements of $\mathcal{A}^*$ belong to $\mathcal{A}$.

We consider divergence measures between fuzzy subsets to try to measure the difference between any two elements in this σ-field formed by fuzzy subsets. To do this, we considered some natural properties of a divergence measure which were that a divergence measure had to be positive and a symmetric function of the two subsets that we are going to compare; it had to be zero when we compared a subset with itself, and it had to be decreasing when the two subsets became more similar. Then we defined a divergence measure like a map D satisfying these three axioms for all fuzzy subsets of the σ-field $\mathcal{A}^*$.

Definition 1. [6] A function $D\colon \mathcal{A}^* \times \mathcal{A}^* \mapsto \mathbb{R}$ is called a *divergence measure between fuzzy subsets* if and only if $\forall A, B \in \mathcal{A}^*$ it satisfies the following conditions:

1. $D(A,B) = D(B,A)$;
2. $D(A,A) = 0$;
3. $\max\{D(A\cup C, B\cup C),\, D(A\cap C, B\cap C)\} \leq D(A,B), \forall C \in \mathcal{A}^*$, where $A\cup B$ and $A\cap B$ are two fuzzy sets defined as $A\cup B(x) = S(A(x),B(x))$ and $A\cap B(x) = T(A(x),B(x)), \forall x \in \Omega$, respectively, for any t-conorm S and t-norm T.

From now on, we are going to consider the t-conorm of the maximum and the t-norm of the minimum, and then this concept is, in the most common cases, a generalization of the concept of distance between two fuzzy subsets.

The most common divergence measures are those called local divergence measures. For these measures, each coordinate is independent from the others and they are all equally important. Formally, we have the following definition:

Definition 2. [6] Let D be a divergence measure over a finite universal set Ω. We say that D is *local* if and only if there exists a function $h\colon [0,1]\times[0,1] \mapsto \mathbb{R}$ satisfying

$$D(A,B) - D(A\cup\Omega^i,\ B\cup\Omega^i) = h(A(x_i),B(x_i)),\ \forall A,B \in \widetilde{\mathcal{P}}(\Omega),\ \forall x_i \in \Omega\,,$$

where

$$\Omega^i(x_j) = \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Note that the local divergence measure can be always defined, since any constant fuzzy subsets belongs to $\mathcal{A}^*$ (see [6]) and therefore, as $\mathcal{A}^*$ is a σ-field, $A\cup\Omega^i \in \mathcal{A}^*,\ \forall A \in \mathcal{A}^*,\ \forall x_i \in \Omega$.

For these special divergence measures the following can be proved [6]:

Proposition 1. *If we have a finite universal set Ω, then D is a local divergence measure if and only if there exists a mapping $h\colon [0,1]\times[0,1] \mapsto \mathbb{R}$ such that*

$$D(A,B) = \sum_{i=1}^{n} h(A(x_i),B(x_i))\,,$$

fulfilling the following conditions:

- $h(x,y) = h(y,x)\,,\ \forall x,y \in [0,1]$;

- $h(x,x) = 0$, $\forall x \in [0,1]$;
- $h(\cdot, y)$ *is a non-increasing function on* $[0,y]$ *and non-decreasing on* $[y,1]$.

Local divergence measures can be considered as the sum of a divergence measure over a singleton applied n times.

Finally, we need a way to evaluate the weight of the fuzzy subsets in the partitions. This can be made through fuzzy measures.

DEFINITION 3. A function $m\colon \mathcal{A}^* \mapsto \mathbb{R}^+$ is called a *fuzzy measure* if and only if $\forall A, B \in \mathcal{A}^*$ it satisfies the following conditions:

1. $m(\emptyset) = 0$;
2. if $A \subseteq B$, then $m(A) \leq m(B)$.

This definition is only an extension of the classical concept of non-additive measure of crisp subsets defined in [11].

Some interesting and frequently used examples of fuzzy measures are these three. The first one is the generalization of the cardinal proposed by De Luca and Termini ([5]). Bouchon-Meunier et al. used the other two in some papers ([2]).

EXAMPLE 1.

1. $m_1(A) = \sum\limits_{x\in\Omega} A(x)$ if Ω is finite.
2. $m_2(A) = \sup\limits_{x\in\Omega} A(x)$.
3. $m_3(A) = \big|\{x \in \Omega\colon A(x) \neq 0\}\big|$ where $|\cdot|$ denotes a measure on Ω.

In this paper, we are going to consider a fuzzy partition. Since there is a lot of definitions of a fuzzy partition, see for instance Bezdek & Harris ([1]), Butnariu ([3]), De Baets & Mesiar ([4]), Ruspini ([9]), Thiele ([12]), etc., we are going to choose the definition of partition that we consider more appropriate in each particular experiment. Thus, a fuzzy partition, in any of above senses, is denoted by $\Pi_A = \{A_i\}_{i=1}^n$ and also by $\Pi_1, \Pi_2, \dots$. We will denote the set of all fuzzy partitions of r elements by $\mathcal{F}_r$.

3. Divergence measures between partitions

Let us now define divergence measures between fuzzy partitions. First, note that it makes sense that divergence measures between fuzzy partitions depend on the divergence between the sets that define the partitions. Besides, the definition should also depend on the measure of the different subsets that define the partitions. Finally, divergence measures between fuzzy partitions should be an extension of divergence measures for fuzzy subsets.

Definition 4. Let D be a divergence measure between sets, let m be a fuzzy measure on $\mathcal{A}^*$ and let r be a natural number. A family of functions $R_r\colon (\mathcal{A}^*)^r \times (\mathcal{A}^*)^r \mapsto \mathbb{R}$ is said to be a *divergence measure between partitions*, if it satisfies that

1. $R_r(\Pi_1, \Pi_2) \geq 0\,,\ \forall\, \Pi_1, \Pi_2 \in (\mathcal{A}^*)^r\,;$
2. $R_r(\Pi_1, \Pi_1) = 0\,,\ \forall\, \Pi_1 \in (\mathcal{A}^*)^r\,;$
3. Let $\Pi_1 = \{A_i\}_{i=1}^r$, $\Pi_2 = \{B_i\}_{i=1}^r$, $\Pi_3 = \{C_i\}_{i=1}^r$ and $\Pi_4 = \{D_i\}_{i=1}^r$ be in $(\mathcal{A}^*)^r$, then there holds
   (a) if for all $i \in \{1, 2, \ldots, r\}$ it holds that $m(A_i) = m(C_i)$, $m(B_i) = m(D_i)$ and $D(A_i, B_i) \geq D(C_i, D_i)$, then
   $$\begin{cases} R_r(\Pi_3, \Pi_4) \leq R_r(\Pi_1, \Pi_2)\,, \\ R_r(\Pi_4, \Pi_3) \leq R_r(\Pi_2, \Pi_1)\,; \end{cases}$$
   (b) if for all $i \in \{1, 2, \ldots, r\}$ it holds that $D(A_i, B_i) = D(C_i, D_i)$ and either $m(A_i) \leq m(C_i) \leq m(D_i) = m(B_i)$ or $m(A_i) \geq m(C_i) \geq m(D_i) = m(B_i)$, then
   $$\begin{cases} R_r(\Pi_3, \Pi_4) \leq R_r(\Pi_1, \Pi_2)\,, \\ R_r(\Pi_4, \Pi_3) \leq R_r(\Pi_2, \Pi_1)\,. \end{cases}$$

We can notice that in the definition of divergence measure between partitions we have not required the symmetry property. This is because sometimes the non-symmetry of divergence measure between partitions could be adequate, like, for instance, when we are going to study experiments in which there is a model or pattern to partition the experience, and we have to compare the different ways of partitioning it with this theoretical model.

When symmetry is verified, we say that the divergence measure between fuzzy partitions is symmetric. Formally,

$$R_r(\Pi_1, \Pi_2) = R_r(\Pi_2, \Pi_1),\ \forall\, \Pi_1,\ \Pi_2 \in (\mathcal{A}^*)^r\,.$$

Usually, we use the symbol R instead of R_r, whenever there is not ambiguity.

R is always referred to divergence measure between partitions, since this is the most interesting case for us, but, in fact, we have given the definition of divergence measures between families of r elements of $\mathcal{A}^*$.

Finally, notice that in the previous definition we have done a matching between the sets of both partitions. Then, how is this matching done?

To answer this question consider the following example:

Example 2. We consider two people, a 55 year-old person and a 25 year-old person. It is obvious that if they have to describe the event of being young,

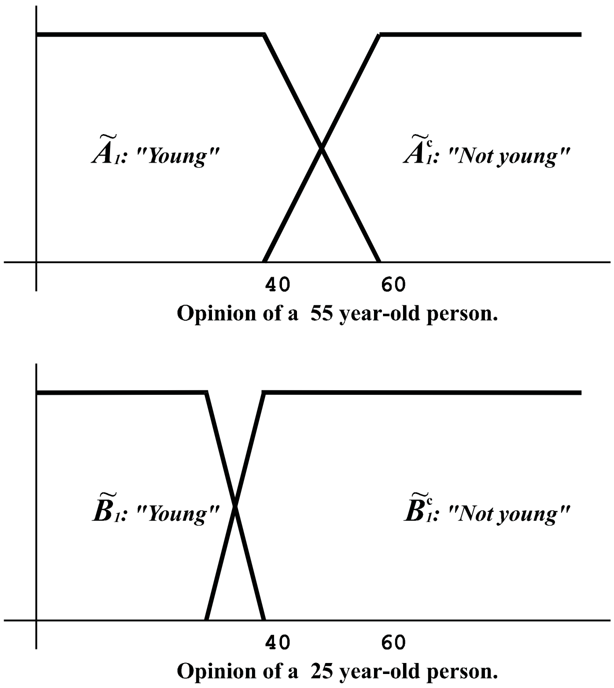

FIGURE 1. Definition of the sets "Young" and its complementary by two different persons.

the fuzzy set, which defines this characteristic, is different for the two people, because one person usually considers as young people those who are close to his age. Suppose that the result of this question is observed in Figure 1.

It is clear that in this case we have to compare the subset A_1 with B_1 and subset A_2 with B_2.

When this comparison order does not exist, or we do not know it, we will apply the Minimum Divergence Principle, that will order the partitions in a way that the sum of the divergences between the subsets is lower than or equal to the sum of these divergences in any other case. This Minimum Divergence Principle is the following:

DEFINITION 5. Let Π_1 and Π_2 be two families in $(\mathcal{A}^*)^r$, formed by the sets $\{A_i\}_{i=1}^r$ and $\{B_i\}_{i=1}^r$, respectively. Let D be a divergence measure between sets. We say that these families are *ordered according to the Minimum Divergence Principle* ($\boldsymbol{D}$), if and only if

$$\sum_{i=1}^{r} D(A_i, B_i) \leq \sum_{i=1}^{r} D\big(A_i, B_{\sigma(i)}\big)\,,$$

for all permutation σ of $\{1, 2, \ldots, r\}$.

Thus, we are going to consider, unless another condition is explicitly indicated, that the two partitions are ordered with this Principle.

4. Local divergence measures

As an example of divergence measures between fuzzy partitions let us see the counterpart of local divergence measures between fuzzy subsets.

Definition 6. Let $R\colon (\mathcal{A}^*)^r \times (\mathcal{A}^*)^r \mapsto \mathbb{R}$ be a divergence measure between partitions associated with (D, m), where r is a natural number. R is said to be a local divergence measure between partitions, if and only if there exists a function g from $\mathbb{R}^3$ in $\mathbb{R}$ such that

$$\begin{aligned} &R\big(\{A_1,\dots,A_i,\dots,A_r\},\{B_1,\dots,B_i,\dots,B_r\}\big) \\ &\qquad - R\big(\{A_1,\dots,A_i,\dots,A_r\},\{B_1,\dots,A_i,\dots,B_r\}\big) \\ &\quad = g\big(D(A_i,B_i),\ m(A_i),\ m(B_i)\big)\,,\ \forall i \in \{1,2,\dots,r\}\,, \\ &\qquad \forall \{A_i\}_{i=1}^r,\ \{B_i\}_{i=1}^r \in (\mathcal{A}^*)^r\,. \end{aligned}$$

The importance of this family of divergence measures lies in the fact that they are very manageable; in fact, the majority of the examples of divergence measures that we are going to see will use this kind of divergences.

As for local divergence measures between fuzzy subsets we can find a characterization of this special class of divergence measures between fuzzy partitions:

Theorem 1. *Consider a divergence measure between partitions R associated with (D, m); then, R is local if and only if for all $\Pi_A, \Pi_B \in (\mathcal{A}^*)^r$ we have that*

$$R(\Pi_A,\Pi_B) = \sum_{i=1}^{r} g\big(D(A_i,B_i), m(A_i), m(B_i)\big)$$

where g satisfies that

1. $g(x,y,z) \ge 0\,,\ \forall (x,y,z) \in G$;
2. $g(0,y,y) = 0\,,\ \forall y \in \mathrm{Im}(m)$;
3. (a) $g(\cdot,y,z)$ *is increasing*, $\forall y,z \in \mathrm{Im}(m)$;
   (b) $g(x,\cdot,z)$ *is decreasing in* $\{y \in \mathbb{R}\colon (x,y,z) \in G \text{ and } y \le z\}$ *and increasing in* $\{y \in \mathbb{R}\colon (x,y,z) \in G \text{ and } y \ge z\}$
   (c) $g(x,y,\cdot)$ *is decreasing in* $\{z \in \mathbb{R}\colon (x,y,z) \in G \text{ and } z \le y\}$ *and increasing in* $\{z \in \mathbb{R}\colon (x,y,z) \in G \text{ and } y \le z\}$

with $G = \big\{(x,y,z) \in \mathbb{R}^3\colon \exists A\ B \in (\mathcal{A}^*)^r \text{ with } x = D(A,B),\ y = m(A),\ z = m(B)\big\}$.

Proof. ($\Rightarrow$) If R is a local divergence measure between partitions, by Definition 6 we have that

$$\begin{aligned} R(\{A_1, A_2, \dots, A_r\}, \{B_1, B_2, \dots, B_r\}) - R(\{A_1, A_2, \dots, A_r\}, \{A_1, B_2, \dots, B_r\}) \\ = g\big(D(A_1, B_1), m(A_1), m(B_1)\big). \end{aligned}$$

By applying this to the partitions $\{A_1, A_2, \dots, A_r\}$ and $\{A_1, B_2, \dots, B_r\}$ we obtain that

$$\begin{aligned} R(\{A_1, A_2, \dots, A_r\}, \{A_1, B_2, \dots, B_r\}) - R(\{A_1, A_2, \dots, A_r\}, \{A_1, A_2, \dots, B_r\}) \\ = g\big(D(A_2, B_2), m(A_2), m(B_2)\big). \end{aligned}$$

Now, following the same process we have that

$$\begin{aligned} R(\{A_1, \dots, A_{r-1}, A_r\}, \{A_1, \dots, A_{r-1}, B_r\}) - R(\{A_1, A_2, \dots, A_r\}, \{A_1, A_2, \dots, A_r\}) \\ = g\big(D(A_r, B_r), m(A_r), m(B_r)\big). \end{aligned}$$

By joining the preceding conclusions and since $R(\{A_i\}_{i=1}^r, \{A_i\}_{i=1}^r) = 0$, we obtain that

$$R(\{A_i\}_{i=1}^r, \{B_i\}_{i=1}^r) = \sum_{i=1}^r g\big(D(A_i, B_i), m(A_i), m(B_i)\big),$$

and therefore, we have only to prove that g satisfies Conditions 1, 2 and 3.

1. Let $(x, y, z) \in G$. Then there exist $A, B \in \mathcal{A}^*$ such that $x = D(A, B)$, $y = m(A)$ and $z = m(B)$. Let $\Pi_A = \{A, A, \dots, A\}$ and $\Pi_B = \{B, B, \dots, B\}$ be two families in $(\mathcal{A}^*)^r$. The divergence measure between them is

$$R(\Pi_A, \Pi_B) = \sum_{i=1}^r g\big(D(A, B), m(A), m(B)\big) = r \cdot g(x, y, z) \geq 0,$$

by applying Axiom 1 in Definition 4.

2. Let y be in $\mathrm{Im}(m)$. Then there exists a fuzzy set $A_1 \in \Omega$ such that $m(A_1) = y$ and hence, $(0, y, y) = \big(D(A_1, A_1), m(A_1), m(A_1)\big) \in G$. Thus we obtain that

$$\begin{aligned} R(\{A_1, A_1, \dots, A_1\}, \{A_1, A_1, \dots, A_1\}) &= \sum_{i=1}^r g\big(D(A_1, A_1), m(A_1), m(A_1)\big) \\ &= r \cdot g(0, y, y) = 0, \end{aligned}$$

by applying Axiom 2 in Definition 4, and hence $g(0, y, y) = 0$.

3. (a) Consider $y, z \in \mathrm{Im}(m)$, and let $x_1, x_2 \in \mathrm{Im}(D)$ be such that $(x_1, y, z), (x_2, y, z) \in G$. Then there exist $A, B, C, D \in \widetilde{\mathcal{P}}(\Omega)$ such

that $x_1 = D(C,D)$ and $x_2 = D(A,B)$, $m(A) = m(C) = y$ and $m(B) = m(D) = z$. Suppose $x_1 \leq x_2$; by considering the families $\Pi_1 = \{A, A, A, \ldots, A\}$, $\Pi_2 = \{B, A, A, \ldots, A\}$, $\Pi_3 = \{C, A, A, \ldots, A\}$ and $\Pi_4 = \{D, A, A, \ldots, A\}$, we obtain that

$$R(\Pi_1, \Pi_2) = g(x_2, y, z) + \sum_{i=2}^{r} g(0, y, y) = g(x_2, y, z),$$

$$R(\Pi_3, \Pi_4) = g(x_1, y, z) + \sum_{i=2}^{r} g(0, y, y) = g(x_1, y, z),$$

but, on the other hand, since the conditions in Axiom 3(a) of Definition 4 are satisfied, we have that $R(\Pi_3, \Pi_4) = g(x_1, y, z) \leq R(\Pi_1, \Pi_2) = g(x_2, y, z)$.

(b) Let $(x, y_1, z), (x, y_2, z)$ be in G; there exist $A, B, C, D \in \widetilde{\mathcal{P}}(\Omega)$ such that $x = D(A,B) = D(C,D), y_1 = m(A)$, $y_2 = m(C)$ and $z = m(B) = m(D)$.

If $y_1 \leq y_2 \leq z$, we have that $m(A) \leq m(C) \leq m(D) = m(B)$ and $D(A,B) = D(C,D)$, therefore if we consider the families in $(\mathcal{A}^*)^r$ defined by $\Pi_1 = \{A, A, \ldots, A\}$, $\Pi_2 = \{B, B, \ldots, B\}$, $\Pi_3 = \{C, C, \ldots, C\}$ and $\Pi_4 = \{D, D, \ldots, D\}$, by applying Axiom 3(b) of Definition 4 we have that

$$R(\Pi_3, \Pi_4) = r \cdot g(x, y_2, z) \leq R(\Pi_1, \Pi_2) = r \cdot g(x, y_1, z).$$

If $y_1 \geq y_2 \geq z$, we have that $m(A) \geq m(C) \geq m(D) = m(B)$ and, by applying again Axiom 3(b), we obtain that $R(\Pi_3, \Pi_4) = r \cdot g(x, y_2, z) \leq R(\Pi_1, \Pi_2) = r \cdot g(x, y_1, z)$; therefore the monotonicity of g in these cases is proved.

(c) There holds either $m(A) = m(C) = y \leq m(D) = z_2 \leq m(B) = z_1$ or $m(A) = m(C) = y \geq m(D) = z_2 \geq m(B) = z_1$ with $D(A,B) = D(C,D) = x$ and analogously to preceding paragraph we obtain that $R(\Pi_3, \Pi_4) = r \cdot g(x, y, z_2) \leq r \cdot g(x, y, z_1) = R(\Pi_1, \Pi_2)$ if $z_1 \leq z_2 \leq y$ or $z_1 \geq z_2 \geq y$.

($\Leftarrow$) Suppose that g satisfies the conditions. Consider

$$R(\Pi_1, \Pi_2) = \sum_{i=1}^{r} g\big(D(A_i, B_i), m(A_i), m(B_i)\big)$$

for all $\Pi_1 = \{A_i\}_{i=1}^r$ and $\Pi_2 = \{B_i\}_{i=1}^r$ in $(\mathcal{A}^*)^r$. Let us prove that R satisfies the axioms in Definition 4:

1. Since the image of g is in $\mathbb{R}^+$, we have that

$$R(\Pi_1, \Pi_2) = \sum_{i=1}^{r} g\big(D(A_i, B_i), m(A_i), m(B_i)\big) \geq 0.$$

2. For all Π_1,

$$\begin{aligned} R(\Pi_1, \Pi_1) &= \sum_{i=1}^{r} g\big(D(A_i, A_i), m(A_i), m(A_i)\big) \\ &= \sum_{i=1}^{r} g\big(0, m(A_i), m(A_i)\big) \\ &= r \cdot 0 = 0\,. \end{aligned}$$

3. Let $\Pi_1 = \{A_i\}_{i=1}^r$, $\Pi_2 = \{B_i\}_{i=1}^r, \Pi_3 = \{C_i\}_{i=1}^r, \Pi_4 = \{D_i\}_{i=1}^r$ be in $(\mathcal{A}^*)^r$,

   (a) If for all $i \in \{1, 2, \ldots, r\}$ we have that $m(A_i) = m(C_i)$, $m(B_i) = m(D_i)$ and $D(A_i, B_i) \geq D(C_i, D_i)$, then

   $$\begin{aligned} R(\Pi_3, \Pi_4) &= \sum_{i=1}^{r} g\big(D(C_i, D_i), m(C_i), m(D_i)\big) \\ &\leq \sum_{i=1}^{r} g\big(D(A_i, B_i), m(A_i), m(B_i)\big) \\ &= R(\Pi_1, \Pi_2)\,; \end{aligned}$$

   and similarly we can prove that $R(\Pi_4, \Pi_3) \leq R(\Pi_2, \Pi_1)$, since g is increasing w.r.t. the first component and D is symmetric.

   (b) If for all $i \in \{1, 2, \ldots, r\}$ we have that $D(A_i, B_i) = D(C_i, D_i)$ and either $m(A_i) \leq m(C_i) \leq m(D_i) = m(B_i)$ or $m(A_i) \geq m(C_i) \geq m(D_i) = m(B_i)$, then

   $$\begin{aligned} R(\Pi_3, \Pi_4) &= \sum_{i=1}^{r} g\big(D(C_i, D_i), m(C_i), m(D_i)\big) \\ &= \sum_{i=1}^{r} g\big(D(A_i, B_i), m(C_i), m(B_i)\big) \\ &\leq \sum_{i=1}^{r} g\big(D(A_i, B_i), m(A_i), m(B_i)\big) \\ &= R(\Pi_1, \Pi_2)\,. \end{aligned}$$

   (c) Analogous.

We have to prove that R is local; let $\Pi_1 = \{A_i\}_{i=1}^r$ and $\Pi_2 = \{B_i\}_{i=1}^r$ be in $(\mathcal{A}^*)^r$, then

$$\begin{aligned}
&R(\{A_1, A_2, \ldots, A_i, \ldots, A_r\}, \{B_1, B_2, \ldots, B_i, \ldots, B_r\}) \\
&\quad - R(\{A_1, A_2, \ldots, A_i, \ldots, A_r\}, \{B_1, B_2, \ldots, A_i, \ldots, B_r\}) \\
&= \left(\sum_{j=1,\, j\neq i}^{r} g\big(D(A_j, B_j),\ m(A_j), m(B_j)\big) + g\big(D(A_i, B_i),\ m(A_i), m(B_i)\big) \right) \\
&= \left(\sum_{j=1,\, j\neq i}^{r} g\big(D(A_j, B_j),\ m(A_j),\ m(B_j)\big) + g\big(D(A_i, A_i),\ m(A_i),\ m(A_i)\big) \right) \\
&= g\big(D(A_i, B_i),\ m(A_i),\ m(B_i)\big) - g\big(D(A_i, A_i),\ m(A_i),\ m(A_i)\big)
\end{aligned}$$

but since $g\big(D(A_i, A_i),\ m(A_i), m(A_i)\big) = g\big(0, m(A_i), m(A_i)\big) = 0$, R is a local divergence between partitions. $\square$

Moreover, if R is symmetric and local, the function g, which characterizes this divergence, satisfies this additional condition, that is, it is symmetric with respect to the second and third variables. Formally,

$$g(x, y, z) = g(x, z, y), \qquad \forall (x, y, z) \in G.$$

5. Relationship with the divergence measure between fuzzy subsets

In this section we are going to relate divergence measures between fuzzy partitions with divergence measures between fuzzy subsets, in the case that Ω is a finite universe.

Theorem 2. *Let $\Omega = \{x_1, x_2, \ldots, x_n\}$ and let R be a divergence measure between partitions associated with D and m, such that*

if $D(A_i, B_i) \geq D(C_i, D_i)$, $\forall i = 1, 2, \ldots, n$, then $R(\Pi_1, \Pi_2) \geq R(\Pi_3, \Pi_4)$,

where $\Pi_1 = \{A_i\}_{i=1}^r$, $\Pi_2 = \{B_i\}_{i=1}^r$, $\Pi_3 = \{C_i\}_{i=1}^r$ and $\Pi_4 = \{D_i\}_{i=1}^r$.

If we define the function $D_R \colon \mathcal{A}^* \times \mathcal{A}^* \mapsto \mathbb{R}$ by

$$D_R(A, B) = R\Big(\big\{\{x_1\}_A,\ \{x_2\}_A, \ldots, \{x_n\}_A\big\},\ \{x_1\}_B, \{x_2\}_B, \ldots, \{x_n\}_B\}\Big),$$

where $\{x_i\}_A$ and $\{x_i\}_B$ are the fuzzy subsets in Ω defined by

$$\{x_i\}_A(x) = \begin{cases} A(x) & \text{if } x = x_i, \\ 0 & \text{otherwise}, \end{cases} \qquad \{x_i\}_B(x) = \begin{cases} B(x) & \text{if } x = x_i, \\ 0 & \text{otherwise}, \end{cases}$$

then D_R is a divergence measure between sets.

Proof. Let A, B and C be in $\mathcal{A}^*$.

1. Since $D(\{x_i\}_A, \{x_i\}_B) = D(\{x_i\}_B, \{x_i\}_A)$, then
$$R\Big(\{\{x_1\}_A, \dots, \{x_n\}_A\},\ \{\{x_1\}_B, \dots, \{x_n\}_B\}\Big)$$
$$= R\Big(\{\{x_1\}_B, \dots, \{x_n\}_B\},\ \{\{x_1\}_A, \dots, \{x_n\}_A\}\Big)$$
and therefore $D_R(A, B) = D_R(B, A)$.

2. We have that $D_R(A, A) = R\Big(\{\{x_1\}_A, \dots, \{x_n\}_A\}, \{\{x_1\}_A, \dots, \dots, \{x_n\}_A\}\Big) = 0$.

3. As $D(\{x_i\}_A \cap \{x_i\}_C, \{x_i\}_B \cap \{x_i\}_C) \le D(\{x_i\}_A, \{x_i\}_B)$, $\forall i = 1, \dots, n$, we have
$$R\Big(\{\{x_1\}_A \cap \{x_1\}_C, \dots, \{x_n\}_A \cap \{x_n\}_C\}, \{\{x_1\}_B \cap \{x_1\}_C, \dots, \{x_n\}_B \cap \{x_n\}_C\}\Big)$$
$$\le R(\{\{x_1\}_A, \dots, \{x_n\}_A\}, \{\{x_1\}_B, \dots, \{x_n\}_B\}),$$
that is, $D_R(A \cap C, B \cap C) \le D_R(A, B)$.

Similarly, we can prove that $D_R(A \cup C, B \cup C) \le D_R(A, B)$. Therefore D_R satisfies the three axioms in Definition 1.

□

We can think that D and D_R are the same divergence measure, but it is not true, as we can see in the following.

EXAMPLE 3. Let Ω be the universe of discourse formed by four elements and let R be the divergence between partitions defined by
$$R(\Pi_A, \Pi_B) = \sum_{i=1}^{4} \big(D(A_i, B_i)\big)^2, \qquad \forall\ \Pi_A = \{A_i\}_{i=1}^4,\ \Pi_B = \{B_i\}_{i=1}^4,$$
where D is the divergence between sets
$$D(A, B) = \sum_{i=1}^{n} \Big[\big|A(x_i) - B(x_i)\big|\Big], \quad \forall A, B.$$

If we consider the sets A and B on Ω, where

Ω	x_1	x_2	x_3	x_4
A	0.3	0.2	0.9	0.5
B	0.2	0.1	0.8	0.3

then $D(A, B) = 0.5$, and however, $D_R(A, B) = 0.07$.

For local divergence measures, the result is:

Corollary 1. *Let $\Omega = \{x_1, \dots, x_n\}$, and let R be a local divergence between partitions associated with a local divergence between sets D and a measure m. If the function g defining R satisfies that*

$$g(x_1, y_1, z_1) \leq g(x_2, y_2, z_2), \qquad \forall x_1, x_2 \in \mathrm{Im}(D),\ x_1 \leq x_2$$

and $m(A) = \frac{\sum A(x_i)}{n}$, $\forall A \in \mathcal{A}^*$, then the function D_R defined in Proposition 2 is a local divergence between sets.

Proof. If for all $x, y \in [0,1]$ we define $h^*(x,y) = g\big(h(x,y), \frac{x}{n}, \frac{y}{n}\big)$, where h is the function defining D, then

$$h^*\big(A(x_i),\ B(x_i)\big) = g\big(D\big(\{x_i\}_A, \{x_i\}_B\big),\ m\big(\{x_i\}_A\big),\ m\big(\{x_i\}_B\big)\big),$$

and therefore $D_R(A,B) = \sum_{i=1}^{n} h^*\big(A(x_i), B(x_i)\big)$, for all A and B in $\mathcal{A}^*$. Consequently, D_R is a divergence between sets and D_R has the local property, since $D_R(A,B) - D_R\big(A \cup \{x_i\},\ B \cup \{x_i\}\big) = h^*\big(A(x_i),\ B(x_i)\big) - h^*(1,1) = h^*\big(A(x_i),\ B(x_i)\big)$. □

6. Relationship with probabilistic divergence measures

Now we are going to define probabilistic divergence measures by means of divergence measures between partitions; for this purpose, the divergences will depend only on the measure (in this case a probability) and have certain additional properties.

Theorem 3. *Let $\Omega = \{x_1, \dots, x_n\}$, and let R be a local divergence between partitions associated with D and a probability p, such that the function g which defines R, as a function of y and z, is convex and $g(x,y,z) = g(0,y,z)$, $\forall x \in \mathbb{R}^+$. If we define*

$$R^*\big(\{p_i\}_{i=1}^n,\ \{q_i\}_{i=1}^n\big) = R\big(\{A_i\}_{i=1}^n,\ \{B_i\}_{i=1}^n\big),$$

where $A_i(x) = p_i$, $B_i(x) = q_i$, $\forall x \in \Omega$, then R^* is a probabilistic divergence measure in the sense of [7].

Proof.

1. $R^*(\{p_i\}_{i=1}^n,\ \{q_i\}_{i=1}^n) = R\big(\{A_i\}_{i=1}^n,\ \{B_i\}_{i=1}^n\big) \geq 0$.
2. $R^*\big(\{p_i\}_{i=1}^n,\ \{p_i\}_{i=1}^n\big) = R\big(\{A_i\}_{i=1}^n,\ \{A_i\}_{i=1}^n\big) = 0$.
3. $R^*\big(\lambda\{p_i\}_{i=1}^n + (1-\lambda)\{p_i'\}_{i=1}^n,\ \{q_i\}_{i=1}^n\big) = \sum_{i=1}^{n} g\big(D\big(\lambda A_i + (1-\lambda)A_i', B_i\big),$ $\lambda p(A_i) + (1-\lambda)p(A_i'),\ p(B_i)\big)$.

Since $g(x,\cdot,z)$ is convex, then

$$
\begin{aligned}
R^*\big(\lambda\{p_i\}_{i=1}^n + (1-\lambda)\{p_i'\}_{i=1}^n,\ \{q_i\}_{i=1}^n\big)&\\
\leq \sum_{i=1}^n \lambda g\big(D\big(\lambda A_i + (1-\lambda)A_i', B_i\big),\ p(A_i),\ p(B_i)\big)&\\
+\sum_{i=1}^n (1-\lambda) g\big(D\big(\lambda A_i + (1-\lambda)A_i', B_i\big),\ p(A_i'),\ p(B_i)\big)&\\
= \sum_{i=1}^n \lambda g\big(D(A_i, B_i), p(A_i),\ p(B_i)\big)&\\
+\sum_{i=1}^n (1-\lambda) g\big(D\big(A_i', B_i\big),\ p(A_i'),\ p(B_i)\big)&\\
= \lambda R^*\big(\{p_i\}_{i=1}^n,\ \{q_i\}_{i=1}^n\big) + (1-\lambda)R^*\big(\{p_i'\}_{i=1}^n,\ \{q_i\}_{i=1}^n\big)&,
\end{aligned}
$$

and therefore R^* is a probabilistic divergence measure. □

7. Some important families of divergence measures between fuzzy partitions

In this section we present three important examples of divergence measures between partitions, each of them having some specific properties. In these three examples we work with local divergences because they are more manageable.

7.1 Divergence measures independent of the measure

The first one is the class formed by divergence measures independent of the measure. This class is important because sometimes it is convenient that the divergence between partitions only depends on the divergence between subsets in the partitions. This divergence measures are defined by

$$R(\Pi_1,\Pi_2) = \sum_{i=1}^r D(A_i, B_i), \qquad \forall\ \Pi_1 = \{A_i\}_{i=1}^r,\ \Pi_2 = \{B_i\}_{i=1}^r \in (\mathcal{A}^*)^r,$$

where D is a divergence between sets.

We can also define these measures as the local divergence measures such that $g(x,y,z) = x\,, \forall (x,y,z) \in G$.

As an example of this class we can consider the Hamming's distance between subsets (see [8]), which is a divergence measure between subsets and thus we obtain a way to measure the difference between two partitions.

7.2 Divergence measures from t-norms

The second example of divergence measure between partitions is the class of the divergence measures based on t-norms. We start with an example.

EXAMPLE 4. If we want to compare the two families represented in Figure 2, it is clear that although the sets A_3 and B_3 are clearly different, the divergence cannot be very large, since the measure (assuming a uniform distribution) of these sets is not very important in relation with the other sets.

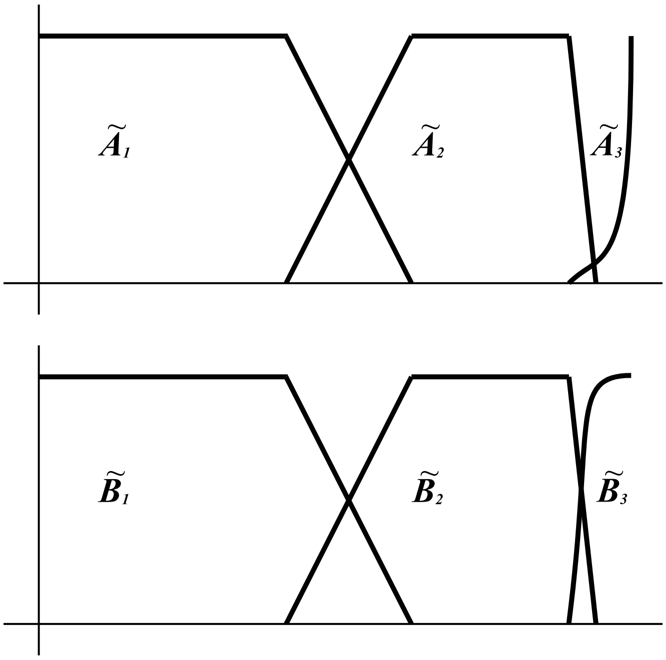

FIGURE 2. Two families in Ω.

Then, we introduce a new way to measure the divergence between two partitions, where the importance of the measure is so high that it will even make the influence on the divergence between two sets to vanish if these sets have a small measure.

This divergence is given by means of the expression

$$R(\Pi_A, \Pi_B) = \sum_{i=1}^{r} \Big[\max\big\{ T\big(D(A_i, B_i)\,,\ m(A_i)\big),\ T\big(D(A_i, B_i),\ m(B_i)\big) \big\} \Big]\,,$$

where T denotes, as usual, a t-norm. Of course, this expression only makes sense if both D and m belong to the interval in which it is defined. Otherwise, we have to normalize D and m to the same interval.

Let us prove that we really have a divergence measure between fuzzy partitions.

Proposition 2. *Let R be defined by*

$$R(\Pi_A, \Pi_B) = \sum_{i=1}^{r} \Big[\max\big\{T\big(D(A_i, B_i),\ m(A_i)\big),\ T\big(D(A_i, B_i),\ m(B_i)\big)\big\}\Big].$$

Then R is a divergence measure between fuzzy partitions.

Proof. We have to prove that

$$g(x, y, z) = \max\big\{T(x, y),\ T(x, z)\big\}$$

satisfies the three conditions in Proposition 1.

1. Trivial.
2. $g(0, x, x) = \max\big\{T(0, x), T(0, x)\big\}$, but since T is a t-norm, then $T(0, u) = 0$, $\forall u \in [0, 1]$, and therefore $g(0, x, x) = \max\{0, 0\} = 0$.
3. (a) If $x_1 \le x_2$, then $T(x_1, y) \le T(x_2, y)$ y $T(x_1, z) \le T(x_2, z)$, and then

$$\begin{aligned} g(x_1, y, z) &= \max\big\{T(x_1, y),\ T(x_1, z)\big\} \le \max\big\{T(x_2, y),\ T(x_2, z)\big\} \\ &= g(x_2, y, z). \end{aligned}$$

(b) If $y_1 \ge y_2 \ge z$, then $T(x, y_1) \ge T(x, y_2) \ge T(x, z)$, and therefore

$$\begin{aligned} & g(x, y_1, z) = \max\big\{T(x, y_1),\ T(x, z)\big\} = T(x, y_1) \\ & \ge T(x, y_2) = \max\big\{T(x, y_2),\ T(x, z)\big\} = g(x, y_2, z). \end{aligned}$$

If $y_1 \le y_2 \le z$, then $T(x, y_1) \le T(x, y_2) \le T(x, z)$, and therefore

$$\begin{aligned} g(x, y_1, z) &= \max\big\{T(x, y_1),\ T(x, z)\big\} = T(x, z) \\ &= \max\big\{T(x, y_2),\ T(x, z)\big\} = g(x, y_2, z). \end{aligned}$$

(c) Trivial by symmetry.

□

Is it possible that R is a local divergence between partitions if we consider another t-conorm different from the maximum? The answer is negative. Let S be a t-conorm, if we consider $g(x, y, z) = S\big(T(x, y), T(x, z)\big)$, and the elements $(M, 0, z)$ and (M, z, z), where $M = \max_{(x,y,z)\in G} \{x\}$ if there exists the maximum, $M = \infty$ otherwise, we obtain that $g(M, 0, z) = S(0, z) = z$, and $g(M, z, z) = S(z, z)$, and therefore, $\forall z \in [0, 1]$ we have that $z \le S(z, z)$ and $z \ge S(z, z)$ ($g(M, \cdot, z)$ is decreasing in $[0, z]$), and therefore we have that $S(z, z) = z$, but the only idempotent t-conorm is the maximum.

Of course, this expression only makes sense if both D and m belong to the interval in which it is defined. Otherwise, we have to normalize D and m to the same interval.

7.3 ϕ-Divergence measures between partitions

The last class of divergence measures between partitions (formed by non null sets) is the divergence constructed from the generalized divergence [7], by defining the divergence as

$$R(\Pi_A, \Pi_B) = \sum_{i=1}^{r} D(A_i, B_i) \cdot m(A_i) \cdot \phi\left(\frac{m(B_i)}{m(A_i)}\right)$$

where ϕ is a function such that

- $\phi(1) = 0$,
- $\phi'(1) = 0$,
- ϕ is convex,
- is twice differentiable.

Proposition 3. *Let R be a function defined by*

$$R(\Pi_A, \Pi_B) = \sum_{i=1}^{r} D(A_i, B_i) \cdot m(A_i) \cdot \phi\left(\frac{m(B_i)}{m(A_i)}\right).$$

Then R is a divergence measure between fuzzy partitions.

Proof. Since $R(\Pi_A, \Pi_B) = \sum_{i=1}^{r} g\big(D(A_i, B_i), m(A_i), m(B_i)\big)$ with

$$g(x, y, z) = x \cdot y \cdot \phi\left(\frac{z}{y}\right),$$

to prove that R is a divergence it suffices to prove that g satisfies the three conditions in Theorem 1.

1. Since ϕ is convex and twice differentiable, ϕ' is an increasing function, and therefore, since $\phi'(1) = 0$, ϕ' is negative in $(-\infty, 1)$ and positive in $(1, \infty)$, and since $\phi(1) = 0$, then $\phi(x) \geq 0$, $\forall x \in \mathbb{R}$, and therefore $g(x, y, z) \geq 0$.
2. $g(0, x, x) = 0 \cdot x \cdot \phi(1) = 0$.
3. (a) If $x_1 \leq x_2$, then $x_1 \cdot y \cdot \phi(\frac{z}{y}) \leq x_2 \cdot y \cdot \phi(\frac{z}{y})$.

   (b) If we consider the function $f_z(y) = y\phi(\frac{z}{y})$, $\forall y \in \mathbb{R}^+$, we have that $f_z'(y) = \phi(\frac{z}{y}) - \frac{z}{y}\phi'(\frac{z}{y})$. Since ϕ is twice differentiable, then it is a continuous function in any interval (a, b) and thus we obtain that $\exists \alpha \in (a, b)$ such that

$$\phi'(\alpha) = \frac{\phi(b) - \phi(a)}{b - a}.$$

• If we apply this theorem in the interval $\left(\frac{z}{y}, 1\right)$ for all $y \geq z$, then

$$\phi\Big(\frac{z}{y}\Big) = \Big(\frac{z}{y} - 1\Big)\phi'(\alpha) = \frac{z}{y}\phi'(\alpha) - \phi'(\alpha), \quad \text{with } \alpha \in \Big(\frac{z}{y}, 1\Big),$$

but since ϕ' is increasing and $\frac{z}{y} \leq \alpha \leq 1$, we have that

$$\phi\Big(\frac{z}{y}\Big) \geq \frac{z}{y}\phi'(\alpha) \geq \frac{z}{y}\phi'\Big(\frac{z}{y}\Big)$$

and therefore, for all $y \geq z$ we have that $f_z'(y) \geq 0$, that is, f_z is increasing, and so, if $y_1 \geq y_2 \geq z$, then $f_z(y_1) \geq f_z(y_2) \Rightarrow g(x, y_1, z) \geq g(x, y_2, z)$ since x is a positive number.

•

$$\phi\Big(\frac{z}{y}\Big) = \Big(\frac{z}{y} - 1\Big)\phi'(\beta) \leq \frac{z}{y}\phi'(\beta)$$

with $\beta \in \left(1, \frac{z}{y}\right)$. Since ϕ' is increasing,

$$\phi\Big(\frac{z}{y}\Big) \leq \frac{z}{y}\phi'\Big(\frac{z}{y}\Big)$$

and therefore f_z is decreasing for all $y < z$; consequently, if $y_1 \leq y_2 \leq z$, then $x \cdot f_z(y_1) \geq x \cdot f_z(y_2)$, and hence $g(x, y_1, z) \geq g(x, y_2, z)$.

(c) Since ϕ is decreasing in $(-\infty, 1]$ and increasing in $[1, \infty)$, then $g(x, y, \cdot)$ is decreasing in $(-\infty, y]$ and increasing in $[y, \infty)$.

□

If we consider that the weight of any x_i is given by the divergence between the sets A_i and B_i, then we obtain a function R that is a divergence measure between partitions with the local property and also a classical probabilistic divergence.

8. Conclusions

In this paper we have introduced and studied the concept of divergence measures between fuzzy partitions. This will allow us to compare two ways to classify a population and to define a fuzziness measure of a fuzzy partition, in the Capocelli and De Luca sense, by comparing a partition with the family formed by the complements of their elements or with the one formed by the closest crisp sets of their elements.

So, we have some immediate open problems, and some other for a future research.

The most immediate one is to generalize the local divergence measure to a larger class, because it is the divergence that we can really manage.

Another open problem is to define a measure of the difference between two fuzzy partitions with a different number of elements. This problem can appear as a simple generalization of the study made in this paper, but it is not true, because in this case we cannot use the preceding works about classical probabilistic divergences, since in this case we also compare partitions with the same number of elements. We think that this is an interesting but non-trivial problem and we hope to be able to solve it in the future.

Finally, we are trying to use all these studies to compare two fuzzy questionnaires with the same number of the goal eventualities, because it was our initial purpose. This is the direction we are following at this time.

REFERENCES

[1] BEZDEK, J. C.—HARRIS, J.: *Fuzzy partitions and relations-an axiomatic basis for clustering*, Fuzzy Sets and Systems **1** (1978), 111–126.

[2] BOUCHON-MEUNIER, B.—RIFQI, M.—BOTHOREL, S.: *Towards general measures of comparison of objects*, Fuzzy Sets and Systems **84** (1996), 143–153.

[3] BUTNARIU, D.: *Additive fuzzy measures and integrals*, J. Math. Anal. Appl. **93** (1983), 436–452.

[4] DE BAETS, B.—MESIAR, R.: *$\mathcal{T}$-partitions*, Fuzzy Sets and Systems **97** (1998), 211–223.

[5] DE LUCA, A.—TERMINI, S.: *A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory*, Inform. and Control **20** (1972), 301–312.

[6] JIMÉNEZ, J.—MONTES, S.—BERTOLUZZA, C.: *Local Divergence Measures on Infinite Referentials*, in: Proceedings of 1999 EUSFLAT-ESTYLF Joint Conference, (1999), pp. 231–234.

[7] KAPUR, J. N.: *Measures of Informations and their Applications*, John Wiley and Sons, New York, Prentice Hall, Englevood Clifs, 1994.

[8] KLIR, G. J.—FOLGER, T. A.: *Fuzzy Sets, Uncertainty, and Information*, John Wiley and Sons, New York, Prentice-Hall International, Englewood Clifs, 1988.

[9] RUSPINI, E. H.: *A new approach to clustering*, Inform. and Control **15** (1969), 22–32.

[10] SALAS, A.—DOLDI, V.—BERTOLUZZA, C.: *Pseudoquestionnaires in a deterministic enviroment*, in: Proceedings of IPMU '2000 Conference, (2000), pp. 1177–1183.

[11] SUGENO, M.: *Theory of Fuzzy Relations and its Applications*, Ph. D. Thesis, Tokyo Institute of Technology, Japan, 1974.

[12] THIELE, H.: *A Characterization of RUSPINI-Partitions by similarity relations*, in: Proceedings of IPMU '96 Conference, (1996), pp. 389–394.

Received May 3, 2000

Susana Montes
Pedro Miranda
Pedro Gil
Department of Statistics and O. R.
University of Oviedo
Campus de Viesques
33271 Gijón
SPAIN
E-mail: smr@pinon.ccu.uniovi.es

Jorge Jiménez
Department of Mathematics
University of Oviedo
Campus de Viesques
33271 Gijón
SPAIN