Mathware & Soft Computing 5 (1998) 253-265

Some Classes of Divergence Measures Between Fuzzy Subsets and Between Fuzzy Partitions*

Susana Montes and Pedro Gil
Departamento de Estadística e I.O. y D.M.
University of Oviedo, C/Calvo Sotelo s/n, 33071 Oviedo, Spain
e-mail: smr,pedro@pinon.ccu.uniovi.es

Abstract

The aim of this paper is to present and study one important class of divergence measure between fuzzy subsets, and one important class of divergence measure between fuzzy partitions, each of them having some specific properties. In the first case, the divergence measure attempts to quantify the degree of difference between two fuzzy subsets $\widetilde{A}$ and $\widetilde{B}$ by comparing the fuzziness of both $\widetilde{A}$ and $\widetilde{B}$ with the fuzziness of the intermediate fuzzy subset. In the second case, we use this divergence between subsets to measure the divergence between partitions.

Keywords: divergence measure, local divergence, fuzziness measure, fuzzy partition.

1 Introduction

Our work regards the study of uncertainty associated with systems in a fuzzy environment. The starting point of our research has been the axiomatic information theory of B.Forte and J.Kampé de Fèriet ([4]), where uncertainty is directly associated with a collection of (crisp) subsets of a space Ω. In the frame of this theory it is possible to guess that there exists a fairly strong relationship between uncertainty (and information) and fuzziness, and between uncertainty and classical divergence. In this respect, a fundamental work has been developed by De Luca-Termini [3], who introduced a kind of measure of fuzziness (the nonprobabilistic entropy of a fuzzy set) based on a probabilistic uncertainty measure.

In our opinion, the bridge between the (crisp) information (with or without probability) and measures of uncertainty and imprecision in fuzzy environments, may lie in what we refer to as divergence, because of the analogy with the classical meaning of the term used by various authors ([9], ...) in comparing two probability distributions. Once we have characterized and studied the measure of the difference

*The research in this paper is supported in part by DGICYT Grant No. PB94-1328.

between two (fuzzy or crisp) subsets, that is, the divergence measure between fuzzy subsets (see [8]), our ultimate objective consists in using this measure to evaluate the fuzziness and/or the uncertainty.

The study of the way to measure the difference between two subsets, to which we will refer to as divergence between subsets, is given in Section 2, and we study mainly the case when the divergence is obtained from fuzziness measures. Departing from this wide class of divergence measure, we study a lot of properties. The way to measure the fuzziness of a fuzzy partitions and the definition of divergence between fuzzy partitions will be given in Section 3. In this section we study also a particular class of divergence between partitions, which is the class of divergences independent of the measure.

2 Divergence measure from fuzziness measure

The measure of the difference of two fuzzy subsets is defined axiomatically on the basis of the following natural properties:

- It is a nonnegative and symmetric function of the two fuzzy subsets to be compared.
- It becomes zero when the two sets coincide.
- It decreases as the two subsets become "more similar" in some sense.

Whereas it is easy to analytically formulate the first an the second condition, the third one depends on the formalization of the concept of "more similar". We base our approach on the fact that if we add (in the sense of union) a subset $\widetilde{C}$ to both fuzzy subsets $\widetilde{A}$ and $\widetilde{B}$, we obtain two subsets which are closer to each other; the same happens with the intersection. So we propose the following

Definition 2.1 *([8]) Let Ω be the universe, and let $\widetilde{P}(\Omega)$ be the family of the fuzzy subset of Ω. A map $D : \widetilde{P}(\Omega) \times \widetilde{P}(\Omega) \longrightarrow I\!R$ is a divergence measure if and only if $\forall \widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega)$, satisfy the following conditions:*

1. $D(\widetilde{A}, \widetilde{B}) = D(\widetilde{B}, \widetilde{A})$;
2. $D(\widetilde{A}, \widetilde{A}) = 0$;
3. $\max\{D(\widetilde{A} \cup \widetilde{C}, \widetilde{B} \cup \widetilde{C}), D(\widetilde{A} \cap \widetilde{C}, \widetilde{B} \cap \widetilde{C})\} \leq D(\widetilde{A}, \widetilde{B}), \forall \widetilde{C} \in \widetilde{P}(\Omega)$, $\forall \widetilde{C} \in \widetilde{P}(\Omega)$.

At the beginning of this section we have indicated that it is natural to assume that the divergence is nonnegative; this condition has not been included in the axioms, since it can be deduced from them.

From now on, we will consider only the case where $\Omega = \{x_1, x_2, \ldots, x_n\}$ is finite. In this case, if we consider the couples $(\widetilde{A}, \widetilde{B}), (\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\})$ which only differ in the i^{th} element (it has been changed from $(\widetilde{A}(x_i), \widetilde{B}(x_i))$ to $(1,1)$), it seems natural to suppose that the variation of divergence only depends on what has been changed. Thus, we introduce the following

Definition 2.2 *([8]) A divergence measure has the local property (or, briefly "is local") if, $\forall \widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega), \forall x_i \in \Omega$, we have that*

$$D(\widetilde{A}, \widetilde{B}) - D(\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\}) = h(\widetilde{A}(x_i), \widetilde{B}(x_i)).$$

This property is natural in general, but not all divergence has the local property. A way to characterize the local divergences is in the following statement.

Proposition 2.3 *([8]) A mapping $D : \widetilde{P}(\Omega) \times \widetilde{P}(\Omega) \longrightarrow \mathbb{R}$, where Ω is a finite frame, $\Omega = \{x_1, x_2, \ldots, x_n\}$ is a local divergence if and only if there exists a function $h : [0,1] \times [0,1] \longrightarrow \mathbb{R}$ such that*

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

and

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

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

iii) $h(x, z) \geq \max\{h(x, y), h(y, z)\}, \forall x, y, z \in [0,1]$ with $x < y < z$.

The preceeding proposition allows us to construct local divergence from a two-side function h. Sometimes some difficulties may arise in verifying condition iii). So we stated the following

Corollary 2.4 *([8]) Condition iii) in Proposition 2.3 can be replaced by iii$'$) $h(\cdot, y)$ is a function decreasing in $[0, y]$ and increasing in $[y, 1]$.*

The first way we choose to compare the membership values is that of comparing the fuzziness of both $\widetilde{A}$ and $\widetilde{B}$ with the fuzziness of the intermediate fuzzy subset. This leads to a wide class of measures. In fact, this was the starting point of our research.

Let us consider the class of fuzziness measures [5] (of local type) given by

$$f(\widetilde{A}) = \sum_{x_i \in \Omega} u(\widetilde{A}(x_i))$$

where $u : [0,1] \longrightarrow I\!R^+$ is a concave function, increasing in $[0,1/2]$, symmetric with respect to the point $1/2$, with $u(0) = u(1) = 0$.

The generator of the fuzziness (function u) can also generate a local divergence. In fact, by posing

$$h(x,y) = u\left(\frac{x+y}{2}\right) - \frac{u(x)+u(y)}{2}, \forall x,y \in [0,1]$$

we obtain a function h which has all the properties required in Proposition 2.3 if u is twice differentiable, so that

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

is a local divergence measure.

To prove this, it is sufficient to prove that h satisfies the conditions i)—iii) in Corollary 2.4. If x, y are in $[0,1]$, we have that:

i) $h(x,y) = u(\frac{x+y}{2}) - \frac{u(x)+u(y)}{2} = u(\frac{y+x}{2}) - \frac{u(y)+u(x)}{2} = h(y,x), \forall x,y \in [0,1]$.

ii) $h(x,x) = u(\frac{x+x}{2}) - \frac{u(x)+u(x)}{2} = u(x) - u(x) = 0$.

iii) We denote by h_y the function $h(\cdot,y)$ for a fixed y. Since h_y is a sum of differentiable functions, it is also differentiable, and hence

$$h_y(x) = u(\frac{x+y}{2}) - \frac{u(x)+u(y)}{2} \Longrightarrow \frac{\partial}{\partial x}h_y(x) = \frac{1}{2}\frac{\partial}{\partial x}u(\frac{x+y}{2}) - \frac{\frac{\partial}{\partial x}u(x)}{2}$$

Since u is concave, we have that $\frac{\partial^2}{\partial x^2}u(x) < 0, \forall x \in [0,1]$, and therefore the function $\frac{\partial}{\partial x}u(x)$ is increasing, whence

$$\begin{cases} x < y \Longrightarrow 2\cdot x < x+y \Longrightarrow \frac{\partial}{\partial x}u(\frac{x+y}{2}) < \frac{\partial}{\partial x}u(x) \\ x > y \Longrightarrow 2\cdot x > x+y \Longrightarrow \frac{\partial}{\partial x}u(\frac{x+y}{2}) > \frac{\partial}{\partial x}u(x) \end{cases}.$$

Thus, $\frac{\partial}{\partial x}h_y(x)$ is negative if $x < y$ and positive if $x > y$; this implies that $h_y(x)$, that is $h(\cdot,y)$, is decreasing in $[0,y]$ and increasing in $[y,1]$. ■

A particular important case of this type is given by the measure obtained from the De Luca-Termini entropy ([3]). Function h obtained from its u function $-x\log x - (1-x)\log(1-x)$ is depicted in Figure 1.

It seems to be evident from the figure that h increases as $|x-y|$ increases, attains its maximum at the points $(0,1)$ and $(1,0)$ ($h(0,1) = h(1,0) = u(\frac{1}{2})$ and its minimum at the points $x = y$ ($h(x,x) = 0$)).

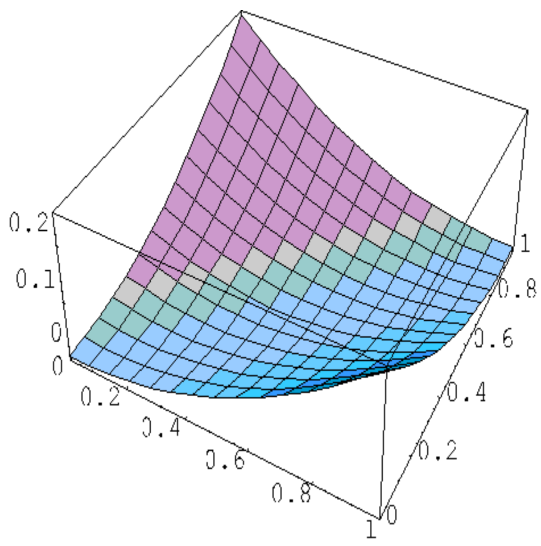

Figure 1: Graphic of $h(x, y)$.

As an example, let us consider the subsets $\widetilde{A}, \widetilde{B}, \widetilde{C}$ (see Table below) of an universe Ω with four elements, defined by (these sets are depicted in Figure 2).

Ω	x_1	x_2	x_3	x_4
$\widetilde{A}$	0.3	0.9	0.6	0.2
$\widetilde{B}$	0.25	0.8	0.6	0.1
$\widetilde{C}$	0.05	0.4	0.9	0.5
$\widetilde{D}$	0.99	0.1	0.1	0.9

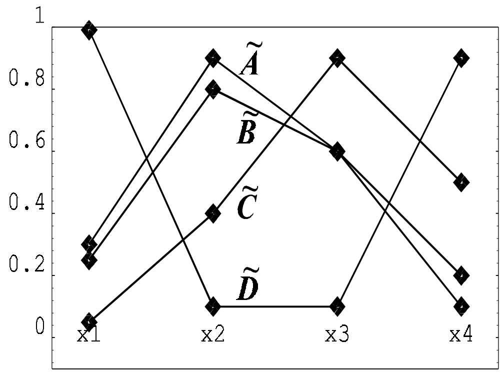

Figure 2: Graphic of $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$ and $\widetilde{D}$.

For these sets we obtain that

$$D(\widetilde{A},\widetilde{B}) = 0.03, D(\widetilde{A},\widetilde{C}) = 0.46, D(\widetilde{A},\widetilde{D}) = 1.60.$$

The divergence measures show that $\widetilde{A}$ is very similar to $\widetilde{B}$ and quite different from $\widetilde{D}$. Notice that the absolute maximum D can assume in this case is given by $D_{max} = 4$.

The divergence based on De Luca-Termini's entropy can be generalized. In fact, in [2] it was proven that a fuzziness measure can be obtained from any uncertainty measure H, provided it satisfies the Principle of Transfer. In this case, we can define $u(t) = H(t, 1-t)$ and the divergence measure takes the form

$$D(\widetilde{A},\widetilde{B}) = \sum_{x\in\Omega}\left[H\left(\left(\frac{\widetilde{A}+\widetilde{B}}{2}\right)(x), \left(\frac{\widetilde{A}+\widetilde{B}}{2}\right)^c(x)\right) - \frac{H(\widetilde{A}(x),\widetilde{A}^c(x)) + H(\widetilde{B}(x),\widetilde{B}^c(x))}{2}\right].$$

It is easy to recognize that, if D is constructed as above from a fuzziness measure, then it can be expressed in terms of function f as follows:

$$D(\widetilde{A},\widetilde{B}) = f(m(\widetilde{A},\widetilde{B})) - \frac{f(\widetilde{A}) + f(\widetilde{B})}{2})$$

where $m(\widetilde{A},\widetilde{B})$ is the "average" of the subsets $\widetilde{A},\widetilde{B}$, that is the fuzzy set defined by

$$m(\widetilde{A},\widetilde{B})(x) = \frac{\widetilde{A}(x)+\widetilde{B}(x)}{2}, \forall x \in \Omega.$$

In the following propositions we establish some important properties of this class of divergence measures, which express natural characteristics of the meaning of our measure. From now on, in this section, we consider that $D(\widetilde{A},\widetilde{B}) = f(m(\widetilde{A},\widetilde{B})) - \frac{f(\widetilde{A})+f(\widetilde{B})}{2}$, where f is any local fuzziness measure.

Proposition 2.5 *Let $\widetilde{A}$ and $\widetilde{B}$ be in $\widetilde{P}(\Omega)$. If $\widetilde{A}$ is sharper than $\widetilde{B}$ ($|\widetilde{A}(x)-1/2| \geq |\widetilde{B}(x)-1/2|, \forall x \in \Omega$), then $D(\widetilde{A},\widetilde{A}^c)$ is greater than or equal to $D(\widetilde{B},\widetilde{B}^c)$, that is,*

$$\text{if } \widetilde{A} \ll \widetilde{B} \text{ then } D(\widetilde{A},\widetilde{A}^c) \geq D(\widetilde{B},\widetilde{B}^c).$$

Proof. Since $f(\widetilde{A}) = f(\widetilde{A}^c) \leq f(\widetilde{B}) = f(\widetilde{B}^c)$ and $\frac{\widetilde{A}+\widetilde{A}^c}{2}(x) = \frac{\widetilde{B}+\widetilde{B}^c}{2}(x) = \frac{1}{2}, \forall x \in \Omega$, then $D(\widetilde{A},\widetilde{A}^c) \geq D(\widetilde{B},\widetilde{B}^c)$. ■

This means, as would be natural, that as the fuzziness decreases, the divergence between a set and its complement increases. It takes the maximum when $\widetilde{A}$ is crisp. Moreover

Proposition 2.6 *Let Z, V be two crisp subsets of Ω, then*

$$D(Z, Z^c) = D(V, V^c).$$

Proof. It is trivial, since $f(Z) = f(Z^c) = f(V) = f(V^c) = 0$. ■

Proposition 2.7 *If we consider the function $f^* : \widetilde{P}(\Omega) \longrightarrow I\!R$ defined by*

$$f^*(\widetilde{A}) = D(Z, Z^c) - D(\widetilde{A}, \widetilde{A}^c), \forall \widetilde{A} \in \widetilde{P}(\Omega),$$

then f^* is a fuzziness measure.

Proof. It is trivial since f^* coincides with f. ■

Proposition 2.8 *Let Z be a crisp subset of Ω. Then $\forall \widetilde{A}, \widetilde{B} \in \Omega$ we have*

$$D(\widetilde{A}, \widetilde{B}) \leq D(Z, Z^c)$$

Proof. It is a consequence that $\frac{f(\widetilde{A})+f(\widetilde{B})}{2} \geq 0$, $f(Z) = f(Z^c) = 0$ and $f(\frac{\widetilde{A}+\widetilde{B}}{2}) \leq f(\widetilde{E}) = f(\frac{Z+Z^c}{2})$, where $\widetilde{E}$ is the equilibrium, that is, the fuzzy subset defined by $\widetilde{E}(x) = 1/2, \forall x \in \Omega$. ■

Proposition 2.9 *The divergence between a fuzzy subsets and the equilibrium is equal to the divergence between its complementary and the equilibrium, that is,*

$$D(\widetilde{A}, \widetilde{E}) = D(\widetilde{A}^c, \widetilde{E}), \forall \widetilde{A}, \widetilde{P}(\Omega).$$

Moreover if $\widetilde{A} \ll \widetilde{B}$ then

$$D(\widetilde{A}, \widetilde{E}) \geq D(\widetilde{B}, \widetilde{E}).$$

Proof. First assertion is a consequence that $u(x) = u(1-x)$ and therefore $u(\frac{x+1/2}{2}) = u(\frac{(1-x)+1/2}{2})$. The second one is a consequence that the properties of monotony and symmetry with respect to $1/2$ of the function $h(x, y) = u\left(\frac{x+y}{2}\right) - \frac{u(x)+u(y)}{2}$ (details can be found in [7]). ■

Although trivial, the following proposition allows us to change the scale factor of a divergence according to our particular requirements

Proposition 2.10 *Let $\phi : I\!R \to I\!R$ be a non decreasing function with $\phi(0) = 0$. The maps D_ϕ, D^ϕ defined below are also divergence measures*

$$D_\phi(\widetilde{A}, \widetilde{B}) = \sum_{x \in \Omega} \phi \left(u \left(\frac{\widetilde{A}(x) + \widetilde{B}(x)}{2} \right) - \frac{u(\widetilde{A}(x)) + u(\widetilde{B}(x))}{2} \right)$$
$$D^\phi(\widetilde{A}, \widetilde{B}) = \phi(D(\widetilde{A}, \widetilde{B}))$$

Proof. For D_ϕ it is only necessary to prove that $\phi\left(u\left(\frac{x+y}{2}\right) - \frac{u(x)+u(y)}{2}\right)$ satisfies the conditions in Proposition 2.3.

The proof to D^ϕ is trivial, since for all $\widetilde{A}, \widetilde{B}$ and $\widetilde{C}$ in $\widetilde{P}(\Omega)$ we have that $\phi(D(\widetilde{A}, \widetilde{B})) = \phi(D(\widetilde{B}, \widetilde{A}))$, $\phi(D(\widetilde{A}, \widetilde{A})) = \phi(0) = 0$, $\phi(D(\widetilde{A} \cup \widetilde{C}, \widetilde{B} \cup \widetilde{C})) \leq \phi(D(\widetilde{A}, \widetilde{B}))$ and $\phi(D(\widetilde{A} \cap \widetilde{C}, \widetilde{B} \cap \widetilde{C})) \leq \phi(D(\widetilde{A}, \widetilde{B}))$ because of D being a divergence measure, $\phi(0) = 0$ and ϕ being not decreasing. ■

Here, we have proven this properties for this particular class of divergence measure obtained from a fuzziness measure, but we can see ([7]) that most of them are true in general for all divergence measure which has the local property.

3 Divergence between partitions independent of the measure

In this section, firstly we study the concept of divergence measure between partitions, after we extend the fuzziness measure to calculate the fuzziness of a fuzzy partition, and finally we present the class of divergence between partitions independent of the measure.

3.1 Divergence measure between fuzzy partitions

It seems logical to think that the divergence between two partitions depends on the divergence between the sets in them, as well as on the measure of these sets. Thus, we have considered that a suitable way adapted of measuring the divergence between two partitions was through a function that depends on the divergence defined in the previous section and on a properly chosen measure of the sets in the partitions.

Similarly to the classical probabilistic divergences, we are going to consider an arbitrary set and to compare two possible ways (say the two partitions established by two expert on the topic) of partitioning it into r fuzzy subsets, by seeing if these two ways are very similar or very different.

We will refer always to divergences between partitions, since this is the most interesting case for us, but in fact we are going to give the definition of divergence between families. Thus, instead of working with the set formed by all partition in Ω, we are going to work with the set formed for the systems of r subsets of Ω, that we will denote by $\widetilde{F}_r$.

Firstly, we have to do some previous comments related to the way to accomplish the matching between the sets of both families, to measure the divergence between them.

If we consider, for instance, the definition of the sets “Young” and “Not Young” of a 25 years-old person and of a 55 years-old person, it is clear that these partitions are not very similar. Suppose that the result of this question is observed in Figure 3.

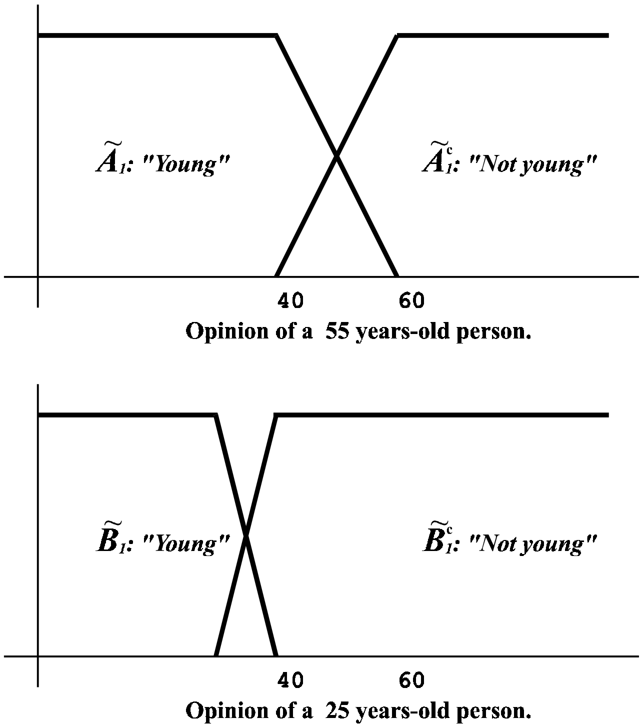

Figure 3: Definition of the sets “Young” and its complementary by two different persons.

It is clear that we have to compare the set $\widetilde{A}_1$ with the set $\widetilde{B}_1$, and $\widetilde{A}_2$ with $\widetilde{B}_2$. When this order to comparing is not fixed, we have to apply the Principle of Minimum Divergence.

Definition 3.1 Principle of Minimum Divergence *([7]) Let $\widetilde{\Pi}_1$ and $\widetilde{\Pi}_2$ be two families in $\widetilde{F}_r$, formed by the sets $\{\widetilde{A}_i\}_{i=1}^r$ and $\{\widetilde{B}_i\}_{i=1}^r$, respectively. Let D be a divergence measure between sets. We say that these families are ordered in accordance with the Principle of Minimum Divergence (D), if and only if for all permutation σ of $\{1, 2, \ldots, r\}$ we have that*

$$\sum_{i=1}^{r} D(\widetilde{A}_i, \widetilde{B}_i) \leq \sum_{i=1}^{r} D(\widetilde{A}_i, \widetilde{B}_{\sigma(i)}), \forall i \in \{1, 2, \ldots, r\}.$$

Thus, in throughout this paper we are going to consider, unless we indicate explicitly another condition, that the two families are ordered with the Principle of Minimum Divergence.

With this criterion, we axiomatize the concept of divergence measure between partitions (in general families).

Definition 3.2 *([7]) Let D be a divergence measure between subsets and let m be a measure on $\widetilde{P}(\Omega)$. A family of functions $R_r : \widetilde{F}_r \times \widetilde{F}_r \longrightarrow I\!R$ is said to be a divergence measure between partitions, if it satisfies that*

1. $R(\widetilde{\Pi}_1, \widetilde{\Pi}_1) \geq 0, \forall \widetilde{\Pi}_1, \widetilde{\Pi}_2 \in \widetilde{F}_r$;
2. $R(\widetilde{\Pi}_1, \widetilde{\Pi}_1) = 0, \forall \widetilde{\Pi}_1 \in \widetilde{F}_r$;
3. *Let* $\widetilde{\Pi}_1 = \{\widetilde{A}_i\}_{i=1}^r$, $\widetilde{\Pi}_2 = \{\widetilde{B}_i\}_{i=1}^r$, $\widetilde{\Pi}_3 = \{\widetilde{C}_i\}_{i=1}^r$ *and* $\widetilde{\Pi}_4 = \{\widetilde{D}_i\}_{i=1}^r$ *be in* $\widetilde{F}_r$, *we have that*

   3.a) If for all $i \in \{1, 2, \ldots, r\}$ *we have that* $m(\widetilde{A}_i) = m(\widetilde{C}_i)$, $m(\widetilde{B}_i) = m(\widetilde{D}_i)$ *and* $D(\widetilde{A}_i, \widetilde{B}_i) \geq D(\widetilde{C}_i, \widetilde{D}_i)$, *then*

   $$\begin{cases} R(\widetilde{\Pi}_3, \widetilde{\Pi}_4) \leq R(\widetilde{\Pi}_1, \widetilde{\Pi}_2) \\ R(\widetilde{\Pi}_4, \widetilde{\Pi}_3) \leq R(\widetilde{\Pi}_2, \widetilde{\Pi}_1) \end{cases};$$

   3.b) If for all $i \in \{1, 2, \ldots, r\}$ *we have that* $D(\widetilde{A}_i, \widetilde{B}_i) = D(\widetilde{C}_i, \widetilde{D}_i)$ *and either* $m(\widetilde{A}_i) \leq m(\widetilde{C}_i) \leq m(\widetilde{D}_i) = m(\widetilde{B}_i)$ *or* $m(\widetilde{A}_i) \geq m(\widetilde{C}_i) \geq m(\widetilde{D}_i) = m(\widetilde{B}_i)$, *then*

   $$\begin{cases} R(\widetilde{\Pi}_3, \widetilde{\Pi}_4) \leq R(\widetilde{\Pi}_1, \widetilde{\Pi}_2) \\ R(\widetilde{\Pi}_4, \widetilde{\Pi}_3) \leq R(\widetilde{\Pi}_2, \widetilde{\Pi}_1) \end{cases}.$$

From now on, we will use the symbol R instead of R_r, whenever there is not ambiguity.

As in Section 2, there exists a very important class of divergence measure between partitions, and this is the class of the local divergences.

Definition 3.3 *([7]) Let $R : \widetilde{F}_r \times \widetilde{F}_r \longrightarrow I\!R$ be a divergence measure between partitions associated with (D, m). R is said to be a local divergence measure between partitions, if and only if the function g from $I\!R^3$ in $I\!R$ satisfies that*

$$R(\{\widetilde{A}_1, \ldots, \widetilde{A}_i, \ldots, \widetilde{A}_r\}, \{\widetilde{B}_1, \ldots, \widetilde{B}_i, \ldots, \widetilde{B}_r\}) -$$

$$-R(\{\widetilde{A}_1, \ldots, \widetilde{A}_i, \ldots, \widetilde{A}_r\}, \{\widetilde{B}_1, \ldots, \widetilde{A}_i, \ldots, \widetilde{B}_r\}) = g(D(\widetilde{A}_i, \widetilde{B}_i), m(\widetilde{A}_i), m(\widetilde{B}_i)),$$

$$\forall i \in \{1, 2, \ldots, r\}, \forall \{\widetilde{A}_i\}_{i=1}^r, \{\widetilde{B}_i\}_{i=1}^r \in \widetilde{F}_r.$$

Though g is defined in $I\!R^3$, we are only interested in the definition of g in the subset of $I\!R^3$ defined by $G = \{(x, y, z) \in I\!R^3 / \exists \widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega)$ with $x = D(\widetilde{A}, \widetilde{B}), y = m(\widetilde{A}), z = m(\widetilde{B})\}$.

The following statement characterizes the local divergences between partitions.

Proposition 3.4 *Let R be a divergence measure between partitions associated with (D, m); R is a local divergence measure between partitions, if and only if for all $\widetilde{\Pi}_{\widetilde{A}}, \widetilde{\Pi}_{\widetilde{B}} \in \widetilde{F}_r$ we have that*

$$R(\widetilde{\Pi}_{\widetilde{A}}, \widetilde{\Pi}_{\widetilde{B}}) = \sum_{i=1}^{r} g(D(\widetilde{A}_i, \widetilde{B}_i), m(\widetilde{A}_i), m(\widetilde{B}_i))$$

where g satisfies that

1. $g(x, y, z) \geq 0, \forall (x, y, z) \in G$;
2. $g(0, y, y) = 0, \forall y \in Im(m)$;
3. *3.a)* $g(\cdot, y, z)$ *is increasing,* $\forall y, z \in Im(m)$;

   3.b) $g(x, \cdot, z)$ *is decreasing in* $\{y \in I\!R / (x, y, z) \in G \text{ e } y \leq z\}$ *and increasing in* $\{y \in I\!R / (x, y, z) \in G \text{ and } y \geq z\}$ *and* $g(x, y, \cdot)$ *is decreasing in* $\{z \in I\!R / (x, y, z) \in G \text{ and } z \leq y\}$ *and increasing in* $\{z \in I\!R / (x, y, z) \in G \text{ and } y \leq z\}$.

3.2 Fuzziness measure of a partition

In this subsection we try to generate fuzziness measures of a fuzzy partition.

Definition 3.5 *A fuzziness measure of a partition is a real function b defined on $\widetilde{F}_r$, such that*

1. $b(\widetilde{\Pi}_1) = 0 \iff \widetilde{A}_i$ *is a crisp set* $\forall \widetilde{A}_i \in \widetilde{\Pi}_1$.
2. *If* $\widetilde{\Pi}_1, \widetilde{\Pi}_2 \in \widetilde{F}_r$ *y* $\Pi_1 \ll \Pi_2$, *then* $b(\widetilde{\Pi}_1) \leq b(\widetilde{\Pi}_2)$.
3. $b(\widetilde{\Pi}_1)$ *is the maximum value between the families with the same number of elements* $\iff \widetilde{A}_i$ *is maximally fuzzy* $\forall \widetilde{A}_i \in \widetilde{\Pi}_1$.

In the second axiom we talk about a partition "sharper than" another one. We have to formalize this concept.

Definition 3.6 *Let $\widetilde{\Pi}_1$ and $\widetilde{\Pi}_2$ be two fuzzy partitions in Ω. The partition $\widetilde{\Pi}_1$ is said to be sharper than $\widetilde{\Pi}_2$, ($\widetilde{\Pi}_1 \ll \widetilde{\Pi}_2$) if and only if*

$$\forall \widetilde{A}_i \in \widetilde{\Pi}_1, \exists \widetilde{B}_{j_i} \in \widetilde{\Pi}_2 / \widetilde{A}_i \text{ is sharper than } \widetilde{B}_{j_i},$$

and

$$\widetilde{B}_{j_i} \neq \widetilde{B}_{j_k}, \forall i \neq k.$$

The following proposition will establish a method to generate fuzziness measures of a partition by means of the well know fuzziness measures of fuzzy subsets.

Proposition 3.7 *Let $f : \widetilde{P}(\Omega) \longrightarrow I\!R$ be a fuzziness measure of fuzzy sets. Then the function $b : \widetilde{F}_r \longrightarrow I\!R$ such that*

$$b(\widetilde{\Pi}_1) = \sum_{i=1}^{r} f(\widetilde{A}_i), \forall \widetilde{\Pi}_1 = \{\widetilde{A}_i\}_{i=1}^{r} \in \widetilde{F}_r$$

is a fuzziness measure of fuzzy partitions.

Proof. It is a consequence of definition of fuzziness measure of fuzzy subsets. ■

3.3 Divergence between partitions independent of the measure

Sometimes, it should be convenient that the divergence between partitions depends only on the divergence between sets in these partitions; an example for this is the following:

$$R(\widetilde{\Pi}_1, \widetilde{\Pi}_2) = \sum_{i=1}^{r} D(\widetilde{A}_i, \widetilde{B}_i), \forall \widetilde{\Pi}_1 = \{\widetilde{A}_i\}_{i=1}^{r}, \widetilde{\Pi}_2 = \{\widetilde{B}_i\}_{i=1}^{r} \in \widetilde{F}_r,$$

where D is a divergence between sets.

To prove that R is a divergence between partitions, it is sufficient to note that R has the local property, and it is defined by $g(x, y, z) = x$; then, by applying Proposition 3.4, it is trivial that R is a local divergence between partitions, since $g(x, y, z) \geq 0, g(0, y, y) = 0, g(x_1, y, z) \leq g(x_2, y, z)$ if $x_1 \leq x_2$ and $g(x, y_1, z) = g(x, y_2, z)$ if $y_1 \leq y_2$ or if $y_1 \geq y_2$, and similar to third component; this is true $\forall (x, y, z) \in I\!R^3$, in particular for all (x, y, z) in G.

Thus, if we consider the divergence between sets defined by means of a fuzziness measure, we obtain that

$$R(\widetilde{\Pi}_1, \widetilde{\Pi}_2) = \sum_{i=1}^{r} f\left(\frac{\widetilde{A}_i + \widetilde{B}_i}{2}\right) - \frac{f(\widetilde{A}_i) + f(\widetilde{B}_i)}{2}$$

but, we can also write this expression as

$$R(\widetilde{\Pi}_1, \widetilde{\Pi}_2) = b\left(\frac{\widetilde{\Pi}_1 + \widetilde{\Pi}_2}{2}\right) - \frac{b(\widetilde{\Pi}_1) + b(\widetilde{\Pi}_2)}{2}$$

where b is the fuzziness measure of fuzzy partitions defined in Proposition 3.7 by means of f.

With this definition, we have proven ([7]) that the crisp partition closest to an ε-partition ([6]) is also the crisp partition for which the divergence is lower, by using the divergence based in the fuzziness measure. We have also proven that the closest ε-partition to a partition in Ruspini's sense ([10]) is the ε-partition with the lowest divergence, when this divergence is measured by the divergence obtained from a fuzziness measure.

Because of the preceding conclusions we can think that if we partitions are those having the lowest distance by using Hamming's distance, then they are have also the lowest divergence. However, this is not true; we can consider $\Omega = \{x_1\}$, and $\widetilde{\Pi}_1 = \{\widetilde{A}_1, \widetilde{A}_2, \widetilde{A}_3, \widetilde{A}_4\}$ with $\widetilde{A}_1(x_1) = 0.4, \widetilde{A}_2(x_1) = 0.3, \widetilde{A}_3(x_1) = 0.2, \widetilde{A}_4(x_1) = 0.1$, $\widetilde{\Pi}_2 = \{\widetilde{B}_1, \widetilde{B}_2, \widetilde{B}_3, \widetilde{B}_4\}$ with $\widetilde{B}_1(x_1) = 1, \widetilde{B}_2(x_1) = 0.2, \widetilde{B}_3(x_1) = 0.2, \widetilde{B}_4(x_1) = 0.1$ and $\widetilde{\Pi}_3 = \{\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_3, \widetilde{C}_4\}$ with $\widetilde{C}_1(x_1) = 0.9, \widetilde{C}_2(x_1) = 0.09$, $\widetilde{C}_3(x_1) = 0.2$, $\widetilde{C}_4(x_1) = 0.1$.

References

[1] A. Bodini, Sugli spazi sfumati di informazione e di incertezza, Tesi de Laurea, Universitá di Pavia, 1995.

[2] I. Couso, P. Gil and S. Montes, Measures of fuzziness and Information Theory, Proceedings of IPMU'96 Conference, 1 (Granada, 1996) 501-505.

[3] A. De Luca and S. Termini, A definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory, Inf. and Control 20 (1972) 301-312.

[4] J. Kampé de Fériet and B. Forte, Information et probabilité, C.R.A.S., 265 (Paris, 1967) 110-114,142-146,350-353.

[5] G.J. Klir and T.A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice-Hall International (UK) Limited, 1988.

[6] S. Montes, J. Jiménez and P. Gil, ε-Partitions and α-Equivalences, Mathware and Soft Computing (1997) to appear.

[7] S. Montes, Partitions and divergence measures in fuzzy models, Thesis, University of Oviedo, 1998.

[8] S. Montes, P. Gil and C. Bertoluzza, Divergence between fuzzy sets and fuzziness, Proceedings of IPMU'98 Conference (Paris, 1998) 1030-1035.

[9] L. Pardo, Teoría de la Información Estadística, Hespérides Salamanca, 1997.

[10] E.H. Ruspini, A new approach to clustering, Inf. Control 15 (1969) 22-32.