ELSEVIER International Journal of Approximate Reasoning 30 (2002) 91–105
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Divergence measure between fuzzy sets

Susana Montes [a,*], Inés Couso [a], Pedro Gil [b], Carlo Bertoluzza [c]

[a] Department of Statistics and O.R., Nautical School, University of Oviedo, 33271 Gijón, Spain
[b] Department of Statistics and O.R., Sciences Faculty, University of Oviedo, 33071 Oviedo, Spain
[c] Department of Computer Sciences and Systems, University of Pavia, Via Ferrata 1, Pavia, Italy

Received 1 February 2001; accepted 1 November 2001

Abstract

In this paper we propose a way of measuring the difference between two fuzzy sets by means of a function which we will call divergence. We define this concept by means of a group of natural axioms and we study in detail the most important classes of such measures, those which have the local property. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: Divergence measure; Local divergence measure; Fuzziness measure; Uncertainty measure

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 Forte [7], where uncertainty is directly associated with a collection of (crisp) subsets of a space Ω (see also [1–3,8–10]). In the frame of this theory it is possible to guess that there exists a fairly strong relationship between uncertainty (and information) and fuzziness. In this respect, a fundamental work has been developed by De Luca–Termini [6], who introduced a

* Corresponding author. Fax: +34 985 182376.
E-mail addresses: smr@pinon.ccu.uniovi.es (S. Montes), couso@pinon.ccu.uniovi.es (I. Couso), pedro@pinon.ccu.uniovi.es (P. Gil), carlo.bertoluzza@unipv.it (C. Bertoluzza).

PII: S0888-613X(02)00063-4

kind of measure of fuzziness (the non-probabilistic entropy of a fuzzy set) which was defined by

$$f(\widetilde{A}) = -\sum_{x\in\Omega}\left[\widetilde{A}(x)\log_2(\widetilde{A}(x)) + \widetilde{A}^c(x)\log_2(\widetilde{A}^c(x))\right]$$

for any $\widetilde{A}$ fuzzy subset of a finite referential Ω.

This measure is based on the probabilistic uncertainty measure proposed by Shannon (Shannon entropy [20]) defined by

$$H(\{p_1,p_2,\ldots,p_n\}) = -\sum_{i=1}^{n}[p_i\log_2(p_i)],$$

where $P = \{p_1,p_2,\ldots,p_n\}$ is a probability measure on Ω.

These previous definitions have been generalized and later these concepts have been axiomatized. In general, a fuzziness measure quantifies the uncertainty concerning our unknowledgement about the inclusion of the elements of Ω in a fuzzy set $\widetilde{A}$, that is, it is a fuzzy uncertainty. On the other hand, an entropy quantifies the uncertainty concerning our unknowledgement about the occurrence of a random experiment, that is, it is a probabilistic uncertainty.

Thus, fuzziness measures and entropies quantify two different kinds of uncertainty. However, we have proven [5] that a fuzziness measure can be obtained from any uncertainty measure H, provided it satisfies the Principle of Transfer. In this case we have that

$$f(\widetilde{A}) = \sum_{x\in\Omega}H(\widetilde{A}(x),\widetilde{A}^c(x)) \quad \forall\widetilde{A}\in\mathscr{P}(\Omega)$$

is a fuzziness measure, where we consider $|\Omega|$ probability systems formed by $\{\widetilde{A}(x),\widetilde{A}^c(x)\}$.

At that time, we suspected that there would exist a strong relationship between probabilistic uncertainty and fuzzy uncertainty, but moreover there exists a strong relationship between probabilistic uncertainty and classical divergence. Thus, we guessed these classical divergence measures could generate some interesting measure in a fuzzy environment. To do this, we were interested in probabilistic divergence measures.

The first one was proposed by Kulback and Leibler [12]. These authors developed the idea from a Jeffreys paper in which the concept of divergence appears to study the problem of finding an invariant density with respect to a probability "a priori". Thus, let $P = \{p_1,p_2,\ldots,p_n\}$, $Q = \{q_1,q_2,\ldots,q_n\}$ be two probability distributions on Ω, Kullback and Leibler quantified the divergence between these two distributions by means of:

$$D(P,Q) = \sum_{i=1}^{n}p_i\log_2\frac{p_i}{q_i}.$$

This divergence measure was later generalized by many authors (see for instance: [4], [18], [21]). Finally, Menéndez et al. [15] tried to propose a generic expression for the most part of the different definitions and they introduce the $(h - \phi)$-divergences.

These definitions were not symmetrical with respect to their arguments, and then the symmetrical version of these divergences was given by means of

$$D^*(P, Q) = D(P, Q) + D(Q, P).$$

Since a probabilistic divergence measure quantifies the difference between two probability distributions, we thought to use these ideas to measure the difference between two fuzzy subsets.

Our initial aim was to define these "fuzzy divergence measures" by means of a general (axiomatic) definition. Then, we intended to use them to define new fuzziness measures as well as to measure the difference between two fuzzy partitions. In the near future, we would like to apply these studies to develop the questionnaire theory in the framework of the fuzzy subsets environment, where we suspect the divergence will play a fundamental role.

The study of the difference between two subsets, to which we will refer to as divergence between subsets, is given in Section 2. Departing from this definition some special class of divergence measures will be studied in Section 3 (the class of local divergence measures). We will study some interesting properties of this wide class. In Section 4 we will propose some examples of divergence measures which are particularly important. These examples provide us divergence measures obtained from fuzziness measures (by using again the link between fuzzy and probabilistic uncertainty), distances (by showing the strong relationship between divergence and distance) and probabilistic divergence measures (in particular we have use the Kullback–Leibler symmetrical divergence). We will conclude presenting some additional comments about this paper and our future researches in this field.

2. Divergence 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 when the two subsets become "more similar" in some sense.

Whereas it is easy to analytically formulate the first and 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 $\widetilde{A}, \widetilde{B}$, we obtain two subsets which are closer to each other; the same happens with the intersection. So we propose the following:

Definition 2.1. Let Ω be the universe we study and let $\widetilde{\mathscr{P}}(\Omega)$ be the family of the fuzzy subset of Ω. A map $D : \widetilde{\mathscr{P}}(\Omega) \times \widetilde{\mathscr{P}}(\Omega) \to \mathbb{R}$ is a divergence measure iff $\forall \widetilde{A}, \widetilde{B} \in \widetilde{\mathscr{P}}(\Omega)$, D satisfies the following conditions:
1. $D(\widetilde{A}, \widetilde{B}) = D(\widetilde{B}, \widetilde{A})$;
2. $D(\widetilde{A}, \widetilde{A}) = 0$;
3. $\max\{D(\widetilde{A} \cup \widetilde{C}, \widetilde{B} \cup \widetilde{C}), D(\widetilde{A} \cap \widetilde{C}, \widetilde{B} \cap \widetilde{C})\} \leqslant D(\widetilde{A}, \widetilde{B}), \forall \widetilde{C} \in \widetilde{\mathscr{P}}(\Omega)$.

Though we can think that the two conditions in Axiom 3 of Definition 2.1 are equivalent, this is not true in general.

The assumption that the divergence is nonnegative can be deduced from Axioms 1 and 2, as follows:

Lemma 2.2. *If D is a divergence measure, then*

$$D(\widetilde{A}, \widetilde{B}) \geqslant 0 \quad \forall \widetilde{A}, \widetilde{B} \in \widetilde{\mathscr{P}}(\Omega).$$

Before continuing with the study of the divergence measures, we are going to examine a first example of divergence.

Example 2.3. Let Ω be a finite universe. The mapping

$$D(\widetilde{A}, \widetilde{B}) = \top_{x \in \Omega} [|\widetilde{A}(x) - \widetilde{B}(x)|],$$

where $\top$ is a *t*-conorm, is a divergence.

It is quite evident that function D satisfies Axioms 1 and 2. The proof of Axiom 3 is more complicated. The details can be found in [16]. We give here a sketch.

We subdivide Ω into the following seven subsets:

$$\begin{aligned} \Omega = & \{x \in \Omega / \widetilde{A}(x) \leqslant \widetilde{B}(x) = \widetilde{C}(x)\} \cup \{x \in \Omega / \widetilde{A}(x) \leqslant \widetilde{B}(x) < \widetilde{C}(x)\} \\ & \cup \{x \in \Omega / \widetilde{A}(x) \leqslant \widetilde{C}(x) < \widetilde{B}(x)\} \\ & \cup \{x \in \Omega / \widetilde{B}(x) < \widetilde{A}(x) \leqslant \widetilde{C}(x)\} \\ & \cup \{x \in \Omega / \widetilde{B}(x) \leqslant \widetilde{C}(x) < \widetilde{A}(x)\} \\ & \cup \{x \in \Omega / \widetilde{C}(x) < \widetilde{A}(x) \leqslant \widetilde{B}(x)\} \\ & \cup \{x \in \Omega / \widetilde{C}(x) < \widetilde{B}(x) < \widetilde{A}(x)\}, \end{aligned}$$

which we will denote by $P_1, \ldots, P_7$. Since $\top$ is associative, we can compute $\top_{x \in \Omega}$ in two steps. Firstly, we compute $\top$ in each of the subsets P_i, then we combine the results, thus obtaining $\top_{x\in\Omega}$. We proved that in each P_i $|(\widetilde{A}\cup\widetilde{C})(x) - (\widetilde{B}\cup\widetilde{C})(x)| \leqslant |\widetilde{A}(x)-\widetilde{B}(x)|$ and $|(\widetilde{A}\cap\widetilde{C})(x) - (\widetilde{B}\cap\widetilde{C})(x)| \leqslant |\widetilde{A}(x)-\widetilde{B}(x)|$. Since $\top$ is monotonic, this suffices to prove that Axiom 3 holds as well.

The following result emphasizes the fact that the closer two sets, the smaller their divergence is.

Proposition 2.4. *Let $\widetilde{A},\widetilde{B},\widetilde{C}$ and $\widetilde{D}$ subsets of Ω such that $\widetilde{A}\subseteq\widetilde{C}\subseteq\widetilde{D}\subseteq\widetilde{B}$. Then $D(\widetilde{C},\widetilde{D})\leqslant D(\widetilde{A},\widetilde{B})$.*

Proof. From $\widetilde{C}\cap\widetilde{D}=\widetilde{C},\widetilde{B}\cap\widetilde{D}=\widetilde{D},\widetilde{C}\cup\widetilde{A}=\widetilde{C},\widetilde{B}\cup\widetilde{C}=\widetilde{B}$, we obtain that $D(\widetilde{C},\widetilde{D})=D(\widetilde{C}\cap\widetilde{D},\widetilde{B}\cap\widetilde{D})\leqslant D(\widetilde{C},\widetilde{B})=D(\widetilde{A}\cup\widetilde{C},\widetilde{B}\cup\widetilde{C})\leqslant D(\widetilde{A},\widetilde{B})$. □

3. Local divergences

In this paragraph we will consider only the case where $\Omega=\{x_1,x_2,\dots,x_n\}$ is finite, and we will denote $a_i=\widetilde{A}(x_i), b_i=\widetilde{B}(x_i), \forall x_i\in\Omega$. Then

$$D(\widetilde{A},\widetilde{B})=F[(a_1,b_1),\dots,(a_n,b_n)].$$

Clearly, F is symmetric in the pairs (a_i,b_i), that is, if $\sigma(1),\dots,\sigma(n)$ is a permutation of $1,\dots,n$, then

$$F[(a_1,b_1),\dots,(a_n,b_n)]=F[(a_{\sigma(1)},b_{\sigma(1)}),\dots,(a_{\sigma(n)},b_{\sigma(n)})].$$

Now let us apply Axiom 3 of Definition 2.1 with $\widetilde{C}=\{x_i\}$. We obtain that

$$F[(a_1,b_1),\dots,(1,1),\dots,(a_n,b_n)]\leqslant F[(a_1,b_1),\dots,(a_i,b_i),\dots,(a_n,b_n)].$$

The pairs $(\widetilde{A},\widetilde{B}),(\widetilde{A}\cup\widetilde{C},\widetilde{B}\cup\widetilde{C})$ only differ in the ith element which has been changed from (a_i,b_i) to $(1,1)$. Thus, it seems natural to suppose that the variation of divergence only depends on what has been changed, that is

$$F[(a_1,b_1),\dots,(a_i,b_i),\dots,(a_n,b_n)]-F[(a_1,b_1),\dots,(1,1),\dots,(a_n,b_n)]=h(a_i,b_i).$$

Thus, we introduce the following:

Definition 3.1. A divergence measure has the local property or, briefly, "is local" if, $\forall\widetilde{A},\widetilde{B}\in\widetilde{\mathscr{P}}(\Omega),\forall x_i\in\Omega$, we have that

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

The following example shows a divergence which does not have the local properties.

Example 3.2. The function

$$D(\widetilde{A},\widetilde{B})=\max_{x\in\Omega}[|\widetilde{A}(x)-\widetilde{B}(x)|]$$

is a divergence, since max is a t-conorm (Example 2.3) but it does not have the local property. In fact, the variation of the divergence values not only depends on the point where the values are changed, but also on the point where $\max[|\widetilde{A}(x) - \widetilde{B}(x)|]$ is located.

This is not the only example of non-local divergence. In fact, it is easy to prove that the divergence defined as in Example 2.3 is always non-local, unless the t-conorm (considered as an associative, commutative, etc, function in any interval (a, b) [19]) is the sum.

Proposition 3.3. *Let Ω be a finite universe of discourse and let $\top$ be a t-conorm, the divergence measure D defined by*

$$D(\widetilde{A}, \widetilde{B}) = \top_{x \in \Omega}[|\widetilde{A}(x) - \widetilde{B}(x)|] \quad \forall \widetilde{A}, \widetilde{B} \in \widetilde{\mathscr{P}}(\Omega)$$

is local if and only if $\top(x, y) = x + y$.

The following statement characterizes the local divergences.

Proposition 3.4. *A mapping $D : \widetilde{\mathscr{P}}(\Omega) \times \widetilde{\mathscr{P}}(\Omega) \to \mathbb{R}$ over a finite frame $\Omega = \{x_1, x_2, \ldots, x_n\}$ is a local divergence iff there exists a function $h : [0, 1] \times [0, 1] \to \mathbb{R}$ 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) $h(x, z) \geqslant \max\{h(x, y), h(y, z)\}, \ \forall x, y, z \in [0, 1]$ *with* $x < y < z$.

Proof. This is a sketch of the proof (details can be found in [16]).

$\Rightarrow$ It is enough to apply Definition 3.1 n times to all the couples (a_i, b_i). At the end of this process we obtain

$$D(\widetilde{A}, \widetilde{B}) = D(\Omega, \Omega)(= 0) + \sum_{i=1}^{n} h(\widetilde{A}(x_i), \widetilde{B}(x_i)).$$

Properties (i)–(ii) are immediate consequences of Axioms 1,2. Property (iii) can be obtained easily by applying Axiom 3 to the three subsets $\widetilde{A}(x_i) = x, \widetilde{B}(x_i) = y, \widetilde{C}(x_i) = z$.

$\Leftarrow$ Properties (i)–(ii) ensure that Axioms 1,2 of Definition 2.1 hold. To obtain Axiom 3 we use the partition of Ω in Example 2.3. We write the sum over Ω, which defines $D(\widetilde{A} \cup \widetilde{C}, B \cup \widetilde{B})$ (or $D(\widetilde{A} \cap \widetilde{C}, B \cap \widetilde{B})$), as a sum of the sums over P_i. In three of the subsets the sum equals zero, in two of them it coincides with the corresponding sum of $D(\widetilde{A}, \widetilde{B})$, and in the remaining two it is lower than or equal to, due to Condition (iii). Thus, Axiom 3 is proved. Finally, it is quite evident that function D is local. □

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

Proposition 3.5. *Condition* (iii) *in Proposition* 3.4 *can be replaced by*:
(iii′) $h(\cdot, y)$ *is a function decreasing in* $[0, y]$ *and increasing* in $[y, 1]$ (*see* [16]).

From Propositions 3.4 and 3.5 we can give two equivalent definitions of local divergence both based on Lemma 3.6. These new definitions will be based in the intersection instead of the union.

Lemma 3.6. *If D is a local divergence on a finite universe Ω, then*

$$D(\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\}) = D(\widetilde{A} \cap \{x_i\}^c, \widetilde{B} \cap \{x_i\}^c) \quad \forall x_i \in \Omega.$$

Proposition 3.7. *A divergence measure D, has the local property iff there exists a function h' such that for all $x_i \in \Omega$,*

$$D(\widetilde{A}, \widetilde{B}) - D(\widetilde{A} \cap \{x_i\}^c, \widetilde{B} \cap \{x_i\}^c) = h'(\widetilde{A}(x_i), \widetilde{B}(x_i)).$$

Proposition 3.8. *A divergence measure D, has the local property iff $\forall \widetilde{A}, \widetilde{B} \in \widetilde{\mathscr{P}}(\Omega), \forall X \in P(\Omega)$*

$$D(\widetilde{A}, \widetilde{B}) = D(\widetilde{A} \cap X, \widetilde{B} \cap X) + D(\widetilde{A} \cap X^c, \widetilde{B} \cap X^c).$$

The lemma and the propositions are stated without proofs, which can be found in [16].

In the following proposition we establish some important properties of the local measures, which express natural characteristics of the meaning of our measure.

Proposition 3.9. *We define on the family $\widetilde{\mathscr{P}}(\Omega)$, a partial ordering* (“*sharper than*”) $\lll$ *by means of* $\widetilde{A} \lll \widetilde{B} \iff |\widetilde{A}(x) - 1/2| \geqslant |\widetilde{B}(x) - 1/2|, \forall x \in \Omega$. *If D has the local property*, *then*

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

Proof. Let $X = \{x \in \Omega / \widetilde{A}(x) \leqslant 1/2\}$ and $Y = \{x \in \Omega / \widetilde{B}(x) \leqslant \widetilde{B}^c(x)\}$.

$$\begin{aligned}D(\widetilde{A},\widetilde{A}^c) &= D(\widetilde{A}\cap X\cap Y,\widetilde{A}^c\cap X\cap Y)+D(\widetilde{A}\cap X\cap Y^c,\widetilde{A}^c\cap X\cap Y^c)\\&\quad+D(\widetilde{A}\cap X^c\cap Y,\widetilde{A}^c\cap X^c\cap Y)+D(\widetilde{A}\cap X^c\cap Y^c,\widetilde{A}^c\cap X^c\cap Y^c)\\&\geqslant D(\widetilde{B}\cap X\cap Y,\widetilde{B}^c\cap X\cap Y)+D(\widetilde{B}\cap X\cap Y^c,\widetilde{B}^c\cap X\cap Y^c)\\&\quad+D(\widetilde{B}\cap X^c\cap Y,\widetilde{B}^c\cap X^c\cap Y)+D(\widetilde{B}\cap X^c\cap Y^c,\widetilde{B}^c\cap X^c\cap Y^c)\\&= D(\widetilde{B},\widetilde{B}^c).\qquad\square\end{aligned}$$

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

Proposition 3.10. *Let Z, V be two crisp subsets of $\Omega=\{x_1,x_2,\ldots,x_n\}$ and let D be a local divergence. Then*

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

Proof. The value of $D(Z,Z^c)$ is $n\cdot h(1,0)$, that is, this value is independent on the elements in Z, and it depends only on the cardinal of the universe, and therefore this divergence coincides for all crisp sets Z in Ω. $\square$

Proposition 3.11. *Let D be a local divergence and let Z be a crisp subset of Ω. Then $\forall\widetilde{A},\widetilde{B}\in\Omega$ we have*

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

Proof. Since D has the local property, then

$$D(\widetilde{A},\widetilde{B})=\sum_{i=1}^{n}h(\widetilde{A}(x_i),\widetilde{B}(x_i))\leqslant n\cdot h(1,0)=D(Z,Z^c)$$

for all $\widetilde{A}$ and $\widetilde{B}$ in $\widetilde{\mathscr{P}}(\Omega)$. $\square$

Proposition 3.12. *Let D be a local divergence over $\Omega=\{x_1,x_2,\ldots,x_n\}$, let c be a generalized fuzzy complement $(\widetilde{A}^c(x_i)=c(\widetilde{A}(x_i)))$ with equilibrium point e, and finally let E_c be the equilibrium set defined by $\widetilde{E}_c(x_i)=e,\forall x_i\in\Omega$. If $h(x,e)=h(c(x),e),\forall x\in\Omega$, then*

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

Proof. It is an immediate consequence of property of the function h. $\square$

If c is the classical complementary suggested by Zadeh ($\widetilde{A}^c(x)=1-\widetilde{A}(x)$, for all $x\in\Omega$), in this proposition we establishes that $h(\cdot,1/2)$ has to be symmetric with respect to $1/2$.

Proposition 3.13. *If* $\forall x_i \in \Omega$ *either* $\widetilde{A}(x_i) \leqslant \widetilde{B}(x_i) \leqslant \widetilde{C}(x_i)$ *or* $\widetilde{C}(x_i) \leqslant \widetilde{B}(x_i) \leqslant \widetilde{A}(x_i)$, *and D is a local divergence, then*

$$D(\widetilde{A}, \widetilde{C}) \geqslant \max\{D(\widetilde{A}, \widetilde{B}), D(\widetilde{B}, \widetilde{C})\}.$$

Proof. It is sufficient to consider the following partition of Ω:

$$\Omega = \{x_i \in \Omega / \widetilde{A}(x_i) \leqslant \widetilde{B}(x_i) \leqslant \widetilde{C}(x_i)\} \cup \{x_i \in \Omega / \widetilde{A}(x_i) \geqslant \widetilde{B}(x_i) \geqslant \widetilde{C}(x_i)\} \cup \{x_i \in \Omega / \widetilde{A}(x_i) = \widetilde{B}(x_i) = \widetilde{C}(x_i)\}. \qquad \square$$

As a consequence, we obtain that if $\widetilde{A}$ is sharper than $\widetilde{B}$, then $D(\widetilde{A}, \widetilde{E}) \geqslant D(\widetilde{B}, \widetilde{E})$.

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

Proposition 3.14. *Let D be a local divergence generated by the function* $h(x, y)$, *and let* $\phi : \mathbb{R} \to \mathbb{R}$ *be a non-decreasing function with* $\phi(0) = 0$. *The maps* D_ϕ *and* D^ϕ *defined below by*

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

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

are also divergence measures and D_ϕ is local.

4. Some classes of divergence

In the subsections of this paragraph we will present and study three important classes of divergence measures, each of them having some specific properties. The divergence measure attempts to quantify the degree of difference between two fuzzy sets $\widetilde{A}, \widetilde{B}$. The local divergence reaches this goal by comparing the membership functions of $\widetilde{A}$ and $\widetilde{B}$ at each point of the reference universe Ω. This can be done in various ways.

4.1. Divergence from fuzziness

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 first application of divergence measures we proposed when we started our research in this field [17].

Let us consider the class of fuzziness measures [11] (of local type) given by

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

where $g : [0,1] \to \mathbb{R}^+$ is a concave function, increasing in $[0, 1/2]$, symmetric with respect to the point $1/2$, with $g(0) = g(1) = 0$. Note that a local fuzziness measure belongs to the class of fuzziness measures proposed by Loo [13] if $F = Id$ and $c_i = 1, f_i = g, \forall i = 1, 2, \ldots, n$ are considered in Loo's definition.

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

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

we obtain a function h which has all the properties required in Proposition 3.4 if g 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.

A particular important case of this type is given by the measure obtained from the De Luca–Termini entropy. Function h obtained from its g function $-x \log x - (1-x)\log(1-x)$ is depicted in Fig. 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 $(1, 0)$ ($h(0,1) = h(1,0) = g(\frac{1}{2})$) and its minimum at the points $x = y$ ($h(x,x) = 0$).

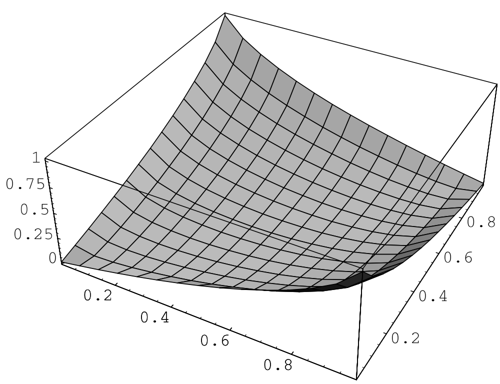

Fig. 1. Graphic of $h(x, y)$.

As an example, let us consider the subsets $\widetilde{A}, \widetilde{B}, \widetilde{C}$ (see Table 1) of an universe Ω with four elements.

For these sets we obtain that

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

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

Of course, De Luca and Termini entropy is only an example. Thus, if we consider a fuzziness measure obtained from an probabilistic entropy H, as we explained in Section 1, we then obtain a particularly important divergence measure, which is defined by means of a probabilistic entropy. The fuzziness measure obtained from H is a local one if we assume $g(t) = H(t, 1-t)$. Thus, the divergence measure assumes 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 fuzzyness measure, then it can be expressed in terms of function f as follows:

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

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

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

This leads us to try to define a fuzzyness directly as

$$D(\widetilde{A}, \widetilde{B}) = \Gamma[f(\widetilde{A}), f(\widetilde{B}), f(m(\widetilde{A}, \widetilde{B}))].$$

Table 1
Membership functions of $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$ and $\widetilde{D}$

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

In particular if, in order to obtain the measure f, we compose the values $g(\widetilde{A}(x_i))$ by means of a strict archimedean conorm in $\mathbb{R}$ (instead of the sum), then $D(\widetilde{A},\widetilde{B})$ as defined above (that is $\Gamma(x,y,z) = z - ((x+y)/2))$ is a divergence, provided that the additive generator θ of the conorm is convex $(\theta(\alpha x + (1-\alpha)y) \leqslant \alpha\theta(x) + (1-\alpha)\theta(y))$.

4.2. Divergence from distance

A particular form of getting local divergence consists in constructing function h by means of a suitable distance in $\mathbb{R}$.

Let d be a distance in $\mathbb{R}$ which also satisfies the following property: if $x < y < z$ then $\max\{d(x,y), d(y,z)\} \leqslant d(x,z)$. This is a natural property, and it is verified by all the most known distances, such as

$$d(x,y) = |x-y| \quad \forall x,y \in \mathbb{R} \qquad \text{Euclidean},$$

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases} \quad \forall x,y \in \mathbb{R} \quad \text{discrete},$$

etc.

Nevertheless, distances exist which do not satisfy the above property, as shown in the following example:

$$d(x,y) = \begin{cases} |x-y| & \text{if } x,y \neq 1,1000 \\ |x-1000| & \text{if } y = 1 \\ |x-1| & \text{if } y = 1000 \\ |1000-y| & \text{if } x = 1 \\ |1-y| & \text{if } x = 1000. \end{cases} .$$

Let ϕ be an increasing (or non-decreasing) function with $\phi(0) = 0$. The function h defined by

$$h(x,y) = \phi(d(x,y)) \quad \forall x,y \in [0,1]$$

satisfies all the properties required in Proposition 3.4. Thus, a local divergence in $\widetilde{\mathscr{P}}(\Omega)$ can be defined by means of

$$D(\widetilde{A},\widetilde{B}) = \sum_{i=1}^{n} \phi[d(\widetilde{A}(x_i),\widetilde{B}(x_i))] \quad \forall \widetilde{A},\widetilde{B} \in \widetilde{\mathscr{P}}(\Omega).$$

If we choose as the distance the Euclidean one, and as the function ϕ the identity, then we measure the divergence by means of the Hamming distance between fuzzy sets

$$d(\widetilde{A},\widetilde{B}) = \sum_{i=1}^{n} |\widetilde{A}(x_i) - \widetilde{B}(x_i)|.$$

This subsection shows that, although divergence and distances are different, it would still be possible to make a confusion in what they measure, but we have to point out that the divergence is more general, and includes the distance as a particular case.

4.3. Entropy-like divergences

The last class of divergence we will propose is related to the Kullbak–Leibler symmetrical probabilistic divergence. It refers to two probability distributions over the same finite space X. In the case where $|X| = 2$ and the two distributions are $\{x, 1-x\}$ and $\{y, 1-y\}$, the Kullbak–Leibler symmetrical function takes on the expression

$$\begin{array}{ll} (x-y)\log_2\left(\frac{y(1-x)}{x(1-y)}\right) & \text{if } x,y \notin \{0,1\}, \\ 0 & \text{if } x = y \in \{0,1\}, \\ \infty & \text{otherwise,} \end{array}$$

which obviously depends on x, y and which we will denote by $h(x, y)$.

From this function we can construct a map $D_J : \widetilde{\mathscr{P}}(\Omega) \times \widetilde{\mathscr{P}}(\Omega) \to \mathbb{R}$ by means of

$$D_J(\widetilde{A}, \widetilde{B}) = \sum_{i=1}^{n} h(\widetilde{A}(x_i), \widetilde{B}(x_i)) \quad \forall \widetilde{A}, \widetilde{B} \in \widetilde{\mathscr{P}}(\Omega),$$

which has all the properties of a local divergence.

It has some relationship with the fuzziness-dependent divergence constructed via the De Luca–Termini fuzziness measure which we will denote by D_{DL}. In particular $D_{DL}(\widetilde{A}, \widetilde{B}) \leqslant D_J(\widetilde{A}, \widetilde{B})$. Following this idea we can construct a lot of divergence measure starting from all the known symmetrical probabilistic divergence measures such as those of order α, type β, class α, β and so on ([14]).

4.4. Concluding remarks

In this paper we propose an axiomatic form to measure the difference between fuzzy sets and we study in detail the case of local divergence. We think that our proposal is quite general and contains, as special cases. almost all the measures known till today. In particular we have completely determined the form of local divergence in the case where the membership function assumes only finite or countable values. We think that this is the main limitation of the present work, so we have tried to extend our study to the continuous case.

The problem we faced in this attempt consists in the generalization of the locality notion. We think that this generalization is not uniquely determinated, but it depends on the choice of a measure m over the range of the membership functions $\widetilde{A}$, which will substitute the sum of the finite and countable case. We have partially studied the case where $\mathrm{ran}(\widetilde{A}) = [0,1]$ and m is the Lebesgue measure, but the general case is still an open problem.

Other open problems regard possible applications of the divergence measure. In particular we think that it could be useful in the detection processes when trying to identify an object (crisp or fuzzy, it does not matter) by means of partially reliable questions. In this case an unreliable answer (naturally represented by a fuzzy set) has to be compared with the set representing the object to be identified. The divergence seems to be the most natural index which measure how they agree.

Acknowledgements

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

References

[1] F. Barbaini, C. Bertoluzza, Sull concetto d'indipendenza in teoria dell'informazione, Statistica XLV (2) (1985).
[2] C. Bertoluzza, M. Schneider, Informations totalement composables, Théories de l'Information, Lect. Notes Math. 398 (1974).
[3] A. Bodini, Sugli spazi sfumati di informazione e di incertezza, Tesi de Laurea, Universitá di Pavia, 1995.
[4] J. Burbea, C.R. Rao, Entropy differential metric distance and divergence measures in probabilistic spaces: a unified approach, Theory Multivariate Anal. 12 (1982) 575–596.
[5] I. Couso, P. Gil, S. Montes, Measures of fuzziness and Information Theory, in: Proceedings of the Conference IPMU'96, Granada, 1996, pp. 501–505.
[6] A. DeLuca, S. Termini, A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory, Inf. Control 20 (1972) 301–302.
[7] B. Forte, Measure of Information: the general axiomatic theory, R.A.I.R.O sèrie R-3 (2) (1969) 63–90.
[8] P. Gil, Medidas de incertidumbre e información en problemas de decisión estadística, Rev. Acad. CC. Ex. Fís. Nat. LXIX (1975) 549–610.
[9] P. Gil, Teoría matemática de la información, ICE, 1981.
[10] J. Kampé de Fériet, P. Benvenuti, Sur une classe d'information, C.R.A.S. 269 (1969) 529–534.
[11] G.J. Klir, T.A. Folger, Fuzzy Sets Uncertainty and Information, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[12] S. Kullback, A. Leibler, On the information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.
[13] S.G. Loo, Measures of fuzziness, Cybernetica 20 (1997) 201–210.
[14] L. Manazza, Misure di entropia, divergenza, imprecisione, Tesi de Laurea, Universitá di Pavia, 1997.

[15] M.L. Menéndez, D. Morales, L. Pardo, M. Salicru, Asymptotic bahaviour and statistical applications of divergence measures in multinomial populations: a unified study, Stat. Papers 36 (1995) 1–29.
[16] S. Montes, Partitions and divergence measures in fuzzy models, Ph.D. Thesis, University of Oviedo, 1998.
[17] S. Montes, P. Gil, Some classes of divergence measures between fuzzy subsets and between fuzzy partitions, Mathware Soft Comput. 5 (3) (1998) 253–265.
[18] M. Salicru, M.L. Menéndez, D. Morales, L. Pardo, A test of independence based on the (r, s)-directed divergence, Tamkang J. Math. 23 (1992) 95–107.
[19] B. Schweizer, A. Sklar, Probabilistic Metric Spaces, North-Holland, New York, 1983.
[20] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.
[21] I.J. Taneja, L. Pardo, D. Morales, M.L. Menéndez, On generalized information and divergence measures and their applications: a brief review, Qestiió 13 (1989) 47–75.