Divergence measures and aggregation operations

P. Miranda
University of Oviedo
Calvo Sotelo s/n, 33007, Oviedo (Spain)
pmm@pinon.ccu.uniovi.es

E. Torres
University of Oviedo
Calvo Sotelo s/n, 33007, Oviedo (Spain)
torres@etsiig.uniovi.es

P. Gil
University of Oviedo
Calvo Sotelo s/n, 33007, Oviedo (Spain)
pedro@pinon.ccu.uniovi.es

Abstract

In this paper we study the relations between divergence measures of two fuzzy sets and other well-known functions, namely aggregation operations and measures of fuzziness. We will show that these functions can be related. Therefore, we study some consequences of these properties as an application; indeed, we obtain some special cases of divergence measures whenever the aggregation operation fulfils some properties. Another application is the translation of some properties of divergence measures to measures of fuzziness.

Keywords: Divergence Measures, Aggregation Operations, Measures of Fuzziness.

1 Introduction and basic concepts

In the following, Ω will denote a set of n elements, $\Omega = \{x_1, ..., x_n\}$. Fuzzy subsets of Ω are denoted by A, B, C and so on. Finally, $\widetilde{\mathcal{P}}(\Omega)$ denotes the set of all fuzzy subsets of Ω.

The comparison of descriptions of objects is a usual operation in many domains: image processing, clustering, deductive reasoning... This comparison is frequently achieved through a measure intended to determine to which extent the descriptions have common points or differ from each other. The measures of comparison have various forms, depending on the purpose of their utilization (see [2] for a classification). A possibility to compare fuzzy subsets are divergence measures.

Divergence measures are used to quantify the difference between two fuzzy subsets of a universal set Ω. They try to maintain the properties of classical divergence measures between two probability distributions which appear in Information Theory. From this concept, it is possible to define divergence measures between fuzzy partitions [13]. We can also use divergence measures to generate measures of fuzziness following the study of difference between the fuzzy subset and its complement [15].

Definition 1. *[13] Let Ω be the universal set and let $\widetilde{\mathcal{P}}(\Omega)$ be the set of all its fuzzy subsets. A mapping $D : \widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \mapsto \mathbb{R}$ is said to be a* **divergence measure between fuzzy subsets** *if and only if the three following axioms hold for any $A, B, C \in \widetilde{\mathcal{P}}(\Omega)$:*

1. $D(A, B) = D(B, A)$.

2. $D(A, A) = 0$.

3. $\max\{D(A \cup C, B \cup C), D(A \cap C, B \cap C)\} \leq D(A, B)$, *in which we have considered the usual definitions based on* max *and* min *for $\cup$ and $\cap$ resp.*

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. *[13] Let $\Omega := \{x_1, ..., x_n\}$ be a finite universal set and 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 $h : [0, 1] \times [0, 1] \mapsto \mathbb{R}$ 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*

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

This definition means that if we make equal a coordinate for both subsets then the change in the divergence only depends on this coordinate.

For these type of divergence measures the following can be proved [13]:

Proposition 1. *D is a local divergence measure over a finite set $\Omega = \{x_1, ..., x_n\}$ if and only if there exists a mapping $h : [0, 1] \times [0, 1] \mapsto \mathbb{R}$ such that*

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

satisfying 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].*

It is easy to prove that these conditions over h are the axioms of divergence measures. Thus, h is a divergence measure over a single referencial set. Then, local divergence measures can be considered as the sum of a divergence measure over n couples of single subsets.

The aggregation operations on fuzzy sets are operations by which several fuzzy subsets are combined to produce a single set. In particular, fuzzy unions and fuzzy intersections are special cases of aggregation operations. There are lots of definitions of aggregation operations, depending on the problem to be solved and even on the author. We will follow the definition given by Klir&Folger in [8].

Definition 3. *An* **aggregation operation** *is a function $s : [0, 1]^n \mapsto [0, 1]$ satisfying*

- *Consider* $\vec{0} = (0, ..., 0)$, $\vec{1} = (1, ..., 1)$. *Then* $s(\vec{0}) = 0, s(\vec{1}) = 1$ *(boundary conditions).*

- *Consider* $\vec{a} = (a_1, ..., a_n)$, $\vec{b} = (b_1, ..., b_n)$. *Then, if* $\vec{a} \leq \vec{b} \Rightarrow s(\vec{a}) \leq s(\vec{b})$ *where* $\vec{a} \leq \vec{b} \Leftrightarrow a_i \leq b_i$ *(monotonicity).*

Measures of fuzziness appear in the seventies to measure the vagueness or fuzziness of a fuzzy set. They have been extensively studied (see [1]).

Definition 4. *[9] A* **measure of fuzziness** *is a mapping*

$$f : \widetilde{\mathcal{P}}(\Omega) \mapsto \mathbb{R}^+$$

satisfying

- $f(A) = 0$ *if and only if* A *is a crisp subset of* Ω.

- *If* $A \preceq B$ *then* $f(A) \leq f(B)$ *where* $A \preceq B$ *is a partial order over* $\widetilde{\mathcal{P}}(\Omega)$ *denoting that* A *is sharper (less fuzzy) than or as fuzzy as* B.

- $f(A)$ *assumes the maximum value if and only if* A *is maximally fuzzy. We will denote the maximally fuzzy subset by* M.

Of course, for each particular conception of the degree of fuzziness, the terms *sharper* and *maximally fuzzy* attain a unique meaning. We will consider $A \preceq B \Leftrightarrow |A(x_i) - A^c(x_i)| \leq |B(x_i) - B^c(x_i)|$, $\forall x_i \in \Omega$, with c a fuzzy complement.

The aim of this paper is to study functional relations between divergence measures and aggregation operations (resp. measures of fuzziness). Our goal is to be able to translate properties of one class to the others. We feel that this can be specially interesting for divergence measures, as aggregation operations and measures of fuzziness are well-known functions. We give an application in Section 2, in which we obtain some bounds for the divergence measure when the aggregation operation used is a t-norm, a t-conorm or an averaging operator. Another possible application (related to Section 3) is the definition of "probabilistic measures of fuzziness", following the same idea as for probabilistic divergence measures developed in [11]. We have already started the research following this way; for the definition of "probabilistic measures of fuzziness" two definitions can be considered: the first one just translates the idea of probabilistic divergence measures; the second one uses the relations obtained in Section 3; it can be proved that these definitions are not the same, but that they coincide for local divergence measures (see [12] for a deeper analysis).

In Section 2 we deal with relations between divergence measures and aggregation operations, while in Section 3 we make the same process for measures of fuzziness and divergence measures. Last section is devoted to conclusions and future research.

2 Relationship between divergence measures and aggregation operations

Let us start giving the general idea. In divergence measures we give a value for the difference of two fuzzy subsets. When we have a finite referencial set, this can be seen as an aggregation of the differences. Now, the question is: How the differences for each coordinate are measured?. The answer is that this is done by some mappings h_i, each of them fulfiling the properties of function

h in Proposition 1. Thus, it makes sense to try to prove that divergence measures are in a sense aggregation operations. This is done in Theorem 1. The reciprocal is given in Theorem 2 in which we build aggregation operations from divergence measures; and a final result joining both results is given in Theorem 3. The section finishes with an application of these results.

Theorem 1. *Let $h_1, ..., h_n$ be n mappings, $h_i : [0,1] \times [0,1] \mapsto [0,1]$, $\forall i$ satisfying:*

1. $h_i(x,x) = 0$, $\forall x \in [0,1]$.
2. $h_i(x,y) = h_i(y,x)$, $\forall x, y \in [0,1]$.
3. $h_i(1,0) = h_i(0,1) = 1$.
4. $h_i(\cdot, y)$ *is a non-increasing function over* $[0,y]$ *and non-decreasing over* $[y,1]$.

Let s be an aggregation operation. Then,

$$D_s(A,B) := ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n))) \tag{2}$$

with $k > 0$, is a divergence measure.

Proof: Let s be an aggregation operation and let us see that D_s defined by (2) is a divergence measure.

1. $D_s(A,B) = D_s(B,A)$ is trivial using condition 2 over h_i.
2. $D_s(A,A) = ks(\vec{0}) = 0$ using condition 1 over h_i.
3. $D_s(A \cup C, B \cup C) = ks(h_1((A \cup C)(x_1), (B \cup C)(x_1)), ..., h_n((A \cup C)(x_n), (B \cup C)(x_n)))$, $D_s(A,B) = ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n)))$. It suffices to prove that $h_i((A \cup C)(x_i), (B \cup C)(x_i)) \leq h_i(A(x_i), B(x_i))$ for all $i = 1, ..., n$. We have three cases:
   - If $C(x_i) \geq A(x_i)$, $C(x_i) \geq B(x_i)$, then $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(C(x_i), C(x_i)) = 0 \leq h_i(A(x_i), B(x_i))$.
   - If $A(x_i) \geq C(x_i) \geq B(x_i)$ or $B(x_i) \geq C(x_i) \geq A(x_i)$, then (considering the first possibility) $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(A(x_i), C(x_i)) \leq h_i(A(x_i), B(x_i))$ because $h_i(x,z) \geq h_i(x,y)$ if $x \geq y \geq z$ (using condition 4 over h_i).
   - If $A(x_i) \geq C(x_i)$, $B(x_i) \geq C(x_i)$, then $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(A(x_i), B(x_i))$.

   The same holds for the intersection.

Then D_s is a divergence measure ■

Let us give some remarks:

1. Condition 3 over h_i is needed so that s can be applied. We could have considered $h_i(x,y) \leq 1$.
2. Divergence measures D_s obtained in the theorem satisfy the following property: If $h_i(A(x_i), B(x_i)) = h_i(C(x_i), D(x_i)), \forall i$, then $D_s(A,B) = D_s(C,D)$ as s is applied over the same vectors.
3. k is needed because divergence measures can reach values greater than 1. If we consider normalized divergence mesures, this constant is not needed at all.

4. Note that the conditions over h_i are the same as those for local divergence measures except for condition 3. This is an important result because it shows that all divergence measures can be generated aggregating divergence measures over single sets. To see this, just remember that function h in local divergence measures was in fact a divergence measure.

Let us consider now two examples:

Example 1. *If we take $h_i = h, \forall i$ and we take the arithmetic mean as aggregation operation we obtain a local divergence measure.*

In fact, note that all local divergence measures can be generated like this, taking $k = n$, as we have

$$D(A,B) = n\frac{\sum_{i=1}^{n} h(A(x_i), B(x_i))}{n} = \sum_{i=1}^{n} h(A(x_i), B(x_i)). \tag{3}$$

Example 2. *Consider $h_i(x,y) = |x-y|, \forall i$. It is trivial to prove that h_i is in the conditions of the theorem. Thus, given s an aggregation operation,*

$$D_s(A,B) := ks(|A(x_1) - B(x_1)|, ..., |A(x_n) - B(x_n)|) \tag{4}$$

with $k > 0$, is a divergence measure.

Note that this result depends on the functions h_i considered, i.e. the divergence measures obtained are different for different choices of h_i.

It is possible to generate divergence measures from other functions instead of aggregation operations. The only condition needed is that if $\vec{p} \leq \vec{q}$ then $s(\vec{p}) \leq s(\vec{q})$ but only for vectors

$$\vec{p}, \vec{q} \in \{(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n)), A, B \in \Omega\} = \mathcal{H}.$$

This allows us to extend Theorem 1:

Proposition 2. *Let $h_1, ..., h_n$ be n mappings, $h_i : [0,1] \times [0,1] \mapsto [0,1]$, $\forall i$ satisfying conditions 1-4 of Theorem 1. Let $s : [0,1]^n \mapsto [0,1]$ be a mapping such that $\vec{p} \leq \vec{q} \Rightarrow s(\vec{p}) \leq s(\vec{q})$ for vectors in $\mathcal{H}$. Then,*

$$D_s(A,B) := ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n))) \tag{5}$$

with $k > 0$, is a divergence measure.

Example 3. *Consider*

$$h_i(x,y) = \begin{cases} 1 & \text{if} \quad \mathrm{x \neq y} \\ 0 & \text{if} \quad \mathrm{x = y} \end{cases}, \forall i$$

Then, s is applied only over 2^n different vectors. For other values, we can consider values such that monotonicity is not verified because these values are never used in the construction of D_s. For example if

$$s(\vec{p}) = \begin{cases} \frac{\sum_{i=1}^{n} p_i}{n} & \text{if} \quad \mathrm{\tilde{p} \in \{0,1\}^n} \\ 1 & \quad \text{otherwise} \end{cases}$$

Then, $D(A,B) = \frac{r}{n}$ with $r \equiv$ number of different coordinates between A, B, is clearly a divergence measure and s is not an aggregation operation.

Proposition 3. *For s in the conditions of Proposition 2 it is always possible to find s' an aggregation operation such that $D_{s'} = D_s$.*

Proof: Define

$$s'(\vec{p}) = sup_{\vec{p'} \leq \vec{p}, \vec{p'} \in \mathcal{H}} s(\vec{p'}) \tag{6}$$

It is easy to see that s' is an aggregation operation and that $s' = s$ over $\mathcal{H}$ ■

For our example, we obtain $s'(\vec{p}) = \frac{r}{n}$ with $r \equiv$ number of coordinates with value 1 in $\vec{p}$.

Of course, we can find other aggregation operators different from s'. In our example, the arithmetic mean leads us to the same divergence measure.

Let us now see the generation of aggregation operation from divergence measures.

Theorem 2. *Let D be a divergence measure. Then,*

$$s_D(\vec{p}) := \frac{D(A, \emptyset)}{D(\Omega, \emptyset)} \tag{7}$$

with $A(x_i) = p_i, \forall i$ is an aggregation operation.

Proof: Let D be a divergence measure and let us see that s_D defined by (7) is an aggregation operation.

- $s(\vec{1}) = \frac{D(\Omega,\emptyset)}{D(\Omega,\emptyset)}$ and thus $s(\vec{1}) = 1$. Now, $s(\vec{0}) = \frac{D(\emptyset,\emptyset)}{D(\Omega,\emptyset)} = 0$.
- Let $\vec{p} \leq \vec{q}$. Then, if we define $A(x_i) = p_i, B(x_i) = q_i$, it suffices to show that $D(B, \emptyset) \geq D(A, \emptyset)$. But this is obvious because $B(x_i) \geq A(x_i)$ and thus, using the third condition of divergence measures we obtain
$$D(B, \emptyset) \geq D(B \cap A, \emptyset \cap A) = D(A, \emptyset). \tag{8}$$

Then s_D is an aggregation operation ■

This aggregation operation does not satisfy necessary unanimity as the following example shows:

Example 4. *Consider*

$$D(A, B) = \begin{cases} 1 & \text{if} \quad \text{A} \neq \text{B} \\ 0 & \text{if} \quad \text{A} = \text{B} \end{cases} \tag{9}$$

It is trivial to see that D is a divergence measure.

Now, if $n = 2$ and we consider $\vec{p} = (0.1, 0.1)$ we obtain $A = (0.1, 0.1)$ and thus $D(A, \emptyset) = 1$, so that $s_D(\vec{p}) = 1$.

However, this result holds sometimes as for example if we consider $D(A, B) = \sum_{i=1}^{n} |A(x_i) - B(x_i)|$.

Finally, we have

Theorem 3. *Let $h_1, ..., h_n$ be n mappings, $h_i : [0, 1] \times [0, 1] \mapsto [0, 1]$ satisfying:*

1. *$h_i(x, x) = 0$, $\forall x \in [0, 1]$.*
2. *$h_i(1, 0) = h_i(0, 1) = 1$.*
3. *$h_i(x, y) = h_i(y, x)$, $\forall x, y \in [0, 1]$.*
4. *$h_i(\cdot, y)$ is a non-increasing function over $[0, y]$ and non-decreasing over $[y, 1]$.*
5. *$Im h_i(., 0) = [0, 1], \forall i$.*

Let D be a non-constant set function satisfying that $D(A,B) = D(C,D)$ whenever $h_i(A(x_i), B(x_i)) = h_i(C(x_i), D(x_i))$ $\forall i$. Then, D is a divergence measure if and only if it can be put as

$$D(A,B) = ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n))) \tag{10}$$

where $k > 0$ and s is an aggregation operation.

Proof: $\Leftarrow$) Given h_i satisfying 1-4 and s an aggregation operation, the set function D given by (10) is a divergence measure by Theorem 1.

$\Rightarrow$) Given h_i satisfying 1-5 and D a divergence measure, let us see that we are able to find s an aggregation operation satisfying (10).

We define $s(\vec{p}) = \frac{D(A,\emptyset)}{D(\Omega,\emptyset)}$ with $h_i(A(x_i), 0) = p_i$.

- s is well-defined: If we have A, B s.t. $h_i(A(x_i), 0) = h_i(B(x_i), 0)$, then $D(A,\emptyset) = D(B,\emptyset)$. We can always find such a subset A by condition 5.
- $s(\vec{1}) = \frac{D(\Omega,\emptyset)}{D(\Omega,\emptyset)}$ and thus $s(\vec{1}) = 1$. Now, $s(\vec{0}) = \frac{D(\emptyset,\emptyset)}{D(\Omega,\emptyset)}$ and then $s(\vec{0}) = 0$.
- Let $\vec{p} \le \vec{q}$. Then, if we define A, B s.t. $h_i(A(x_i), 0) = p_i, h_i(B(x_i), 0) = q_i$, it suffices to show that $D(B,\emptyset) \ge D(A,\emptyset)$. But this is obvious because $B(x_i) \ge A(x_i)$ by condition 4 and thus, using the third condition of divergence measures we obtain

$$D(B,\emptyset) \ge D(B \cap A, \emptyset \cap A) = D(A,\emptyset). \tag{11}$$

- Finally, we have to prove that $D = ks$. Consider $k = D(\Omega,\emptyset)$.

  Define $\vec{p} = (h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n)))$.

  We define C s.t. $h_i(C(x_i), 0) = h_i(A(x_i), B(x_i))$. This is always possible by condition 5. Then,

  $D(A,B) = D(C,\emptyset) = D(\Omega,\emptyset)s(\vec{p}) = ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n)))$ ■

Remark 1. *It might not be possible to find s for each D and h_i. To see this, consider $|\Omega| = 1, D(A,B) = |A(x) - B(x)|$ and*

$$h(x,y) = \begin{cases} 1 & \text{if} \quad \text{x} \neq \text{y} \\ 0 & \text{if} \quad \text{x} = \text{y} \end{cases}$$

Then, D can take infinite many values and h can take only two. Thus, it is not possible to find an aggregation operation relating them because there are different pairs of subsets $(A,B), (C,D)$ such that $h_i(A(x_i), B(x_i)) = h_i(C(x_i), D(x_i)), \forall i$ but $D(A,B) \neq D(C,D)$. Thus condition 5 is needed.

From these results we can try to translate properties from a family to the other one. For example, if s is a continuous function we can study whether D_s is a continuous function (as a real function of $2n$ variables). For the general case the result is not true because the mappings h_i are not necessarily continuous. However, the result holds if h_i is continuous for all i as D would be a composition of continuous functions.

Another question is the following: If s is a symmetric function, can we assure that D_s is a symmetric divergence measure?. The answer is again negative; the reason is that we can not assure that all functions h_i are equal. However, if $h_i = h_j, \forall i, j$, then D_s is symmetric.

In [8], three classes of aggregation operations that are specially important are given: t-norms, t-conorms and averaging operators. Let us now see some properties about each class:

Proposition 4.

- *Let s be a t-norm. Then*

$$D_s(A,B) \leq kmin(h_1(A(x_1),B(x_1)),...,h_n(A(x_n),B(x_n))). \tag{12}$$

In particular, if $\exists i$ *s.t.* $h_i(A(x_i),B(x_i)) = 0$*, then* $D_s(A,B) = 0$*.*

- *Let s be a t-conorm. Then,*

$$D_s(A,B) \geq kmax(h_1(A(x_1),B(x_1)),...,h_n(A(x_n),B(x_n))). \tag{13}$$

In particular, if $\exists i$ *s.t.* $h_i(A(x_i),B(x_i)) = 1$*, then* $D_s(A,B) = k$ *with k the maximal value reached by* D_s*.*

- *If s is an averaging operator, then*

$$D_s(A,B) \in k[min_i(h_i(A(x_i),B(x_i))), max_i(h_i(A(x_i),B(x_i)))]. \tag{14}$$

In particular, if $h_i(A(x_i),B(x_i)) = constant, \forall i$*, then* $D(A,B) = k\,constant$*.*

Proof: Trivial ■

This result shows that when a divergence measure comes from a t-norm, then if A, B are such that they are very similar for one coordinate (in terms of h_i), then $D_s(A,B)$ will be also very small, no matter how different the other coordinates are.

Note that in this case we can consider divergence measures as homotecies of t-norms. However, the interpretation is completely different: for t-norms, we have n fuzzy subsets and only one element; however, for divergence measures we have two fuzzy subsets and n elements.

The same holds for t-conorms, i.e. if a divergence measure comes from a t-conorm, then if A, B are such that they are very different for one coordinate (in terms again of h_i), then $D_s(A,B)$ will be also very big, no matter how similar the other coordinates are.

Finally, for averaging operators we obtain a similar result, and we can also see that in this case we obtain idempotent divergence measures.

3 Relationship between divergence measures and measures of fuzziness

Let us now make the same process for divergence measures and measures of fuzziness. This section comes up from the relations between aggregation operations and measures of fuzziness obtained in [10]. Thus, this section completes the relations among these three measures. We will follow the same pattern as in Section 2. Thus, Theorem 4 studies the generation of measures of fuzziness from divergence measures, Theorem 5 deals with the reverse result and in Theorem 6 we join these two results. The section finishes with a study of probabilistic measures of fuzziness.

We start generating measures of fuzziness from divergence measures.

Theorem 4. *Consider the classical fuzzy union and fuzzy intersection proposed by Zadeh in [17]. Let c be an involutive fuzzy complement. Let D be a divergence measure satisfying:*

- $D(A, A^c)$ *attains its maximal value if and only if A crisp (i.e., for all crisp subset Z,* $D(Z, Z^c)$ *is maximum and it is the same for all Z).*

- *$D(A,B) = 0 \Leftrightarrow A = B$.*
- *$D(A,B)$ only depends on $|A(x) - B(x)|$, $\forall x \in \Omega$, i.e., if A, B, C, D are such that $|A(x) - B(x)| = |C(x) - D(x)|$, $\forall x \in \Omega$, then $D(A,B) = D(C,D)$.*

Then,

$$f_D(A) = D(Z, Z^c) - D(A, A^c), \tag{15}$$

with Z a crisp subset is a measure of fuzziness.

Proof: Let D be as in the theorem and let us prove that f_D is a measure of fuzziness.

- If A is crisp, $f_D(A) = 0$ by hypothesis 1 and these are the only subsets satisfying this condition.
- Let $A \preceq B$. Note that $D(A, A^c) = D(A', A'^c)$ with

$$A'(x) = \begin{cases} A(x) & if \quad A(x) \leq e \text{ (the equilibrium point)} \\ A^c(x) & if \quad A(x) > e \end{cases}$$

  Now, we have $D(A', A'^c) \geq D(A' \cup B', A'^c \cup B') = D(B', A'^c) \geq D(B' \cap B'^c, A'^c \cap B'^c) = D(B', B'^c) = D(B, B^c)$, and thus $f_D(A) \leq f_D(B)$.
- If M is the maximally fuzzy subset, we have $M = M^c$ and thus $D(M, M^c) = 0$. Then, f takes its maximum at M. Note that this is the only maximum by hypothesis 2 ■

Let us make some remarks:

- Note that f_D verifies $f(A) = f(A^c)$. This property is considered an axiom by many authors ([5]).
- We need an involutive fuzzy complement in order to have an equilibrium point (necessary in the definition of M).

Let us now turn to the generation of divergence measures from measures of fuzziness:

Theorem 5. *Consider the classical fuzzy union and fuzzy intersection proposed by Zadeh. Let c be an involutive fuzzy complement. Let f be a measure of fuzziness such that $f(A) = f(A^c)$. Then,*

$$D_f(A,B) = f(M) - f(C_{A,B}) \tag{16}$$

with $C_{A,B}(x_i)$ s.t. $c(C_{A,B}(x_i)) - C_{A,B}(x_i) = |A(x_i) - B(x_i)|$ is a divergence measure.

Proof: Let f be a measure of fuzziness. Let us see that given A, B, there exists C such that $D(C, C^c) = D(A,B)$. As c is a continuous complement (c is involutive), the mapping $c - c^2 = c - Id$ is also a continuous function. Now, $c(0) - 0 = 1, c(1) - 1 = -1$. Then, for each $x_i \in \Omega$ there exists v_i such that $c(v_i) - v_i = |A(x) - B(x)|$. This value of v_i is unique: Suppose that there exists another value $v_i' > v_i$ satisfying $c(v_i') - v_i' = |A(x) - B(x)|$. Then, as c is a non-increasing function, $c(v_i') \leq c(v_i)$ and thus $c(v_i') - v_i' < c(v_i) - v_i$ contradicting our hypothesis. We define $C_{A,B}(x_i) = v_i$. Now, let us see that D_f is a divergence measure.

1. As $|A(x) - A(x)| = 0$, then $C_{A,A}(x) = e$, the equilibrium point of c (this point always exists and is unique for continuous complements). Thus, $C_{A,A} = M$ and we obtain $D_f(A,A) = 0, \forall A \in \mathcal{P}(\Omega)$.

2. $D(A,B) = D(B,A)$ is trivial.

3. The third condition is trivial too. Just note that $C_{A,B}$ is sharper than $C_{A\cup D,B\cup D}$ and thus $f(C_{A\cup D,B\cup D}) \geq f(C_{A,B})$ and consequently $D_f(A,B) \geq D_f(A\cup D, B\cup D)$. The same holds for the intersection.

Then D_f is a divergence measure ■

It is easy to see that D_f satisfies:

- $D(A,A^c)$ attains its maximal value if and only if A crisp.
- $D(A,B) = 0 \Leftrightarrow A = B$.
- $D(A,B)$ only depends on $|A(x)-B(x)|$, $\forall x \in \Omega$, i.e., if A,B,C,D are such that $|A(x)-B(x)| = |C(x)-D(x)|$, $\forall x \in \Omega$, then $D(A,B) = D(C,D)$.

Example 5. *Consider the measure of fuzziness*

$$f(A) = \sum_{i=1}^{n} |A(x_i) - C(x_i)| \tag{17}$$

where C is the nearest crisp set in terms of the Hamming distance. This measure of fuzziness belongs to a wider class of measures defined by Kaufmann in [7]. If we apply the previous result with $c(x) = 1 - x$ we obtain

$$D(A,B) = n - f(C) \tag{18}$$

with C such that $C(x) - C^c(x) = |A(x) - B(x)|, \forall x \in \Omega$.

If we join these two results, we obtain:

Theorem 6. *Consider the classical fuzzy union and fuzzy intersection proposed by Zadeh. Let c be an involutive fuzzy complement. Let f be a mapping over fuzzy subsets such that $f(A) = f(A^c)$. Then, f is a measure of fuzziness if and only if f can be put as*

$$f(A) = D(Z,Z^c) - D(A,A^c) \tag{19}$$

with Z a crisp subset and D a divergence measure satisfying

1. *$D(A,A^c)$ attains its maximum value if and only if A crisp.*
2. *$D(A,B) = 0 \Leftrightarrow A = B$.*
3. *$D(A,B)$ only depends on $|A(x)-B(x)|$, $\forall x \in \Omega$, i.e., if A,B,C,D are such that $|A(x)-B(x)| = |C(x)-D(x)|$, $\forall x \in \Omega$, then $D(A,B) = D(C,D)$.*

Proof:
$\Leftarrow$) Define $D(A,B) = f(M) - f(C_{A,B}$ as in Theorem 5. Then, D is a divergence measure. Besides, $f(M) = D(\Omega,\emptyset)$, $D(A,A^c) = f(M) - f(A)$ and thus $f(A) = D(\Omega,\emptyset) - D(A,A^c)$.
$\Rightarrow$) Theorem 4 ■

Let us see one final remark:

Remark 2. *Here, we have considered* $A \preceq B \Leftrightarrow |A(x) - A^c(x)| \leq |B(x) - B^c(x)|, \forall x \in \Omega$, *with* c *a fuzzy complement. We can adapt the result for other definitions of "less fuzzy than". We have had considered* $|A(x_i) - B(x_i)|$ *because this is the usual definition for this concept.*

These are not the only possible relations. For local measures, there are other possibilities for relating measures of fuzziness and divergence measures (see [14]).

A possible application related to these results is the definition of "probabilistic measures of fuzziness". In [11] we have defined the probabilistic divergence measures as

Definition 5. *Let* D *be a divergence measure and* P *representing a finite discrete distribution over* Ω. *We define the* **probabilistic divergence measure** *(or probabilistic divergence) associated to* D *and* P *as:*

$$D_P(A, B) = \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} (p_{(i)} - p_{(i-1)}) D(A_{(i)}, B_{(i)}), \tag{20}$$

where parentheses for probabilities mean a permutation such that $0 = p_{(0)} \leq p_{(1)} \leq ... \leq p_{(n)}$ *and* $A_{(i)}(x_j) = min\{A(x_j), \Omega_i(x_j)\}$ *with*

$$\Omega_i(x_j) = \begin{cases} 1 & if \quad p(x_j) \geq p_{(i)} \\ 0 & if \quad p(x_j) < p_{(i)} \end{cases}$$

and the same holds for $B_{(i)}$. $\mathcal{E}(P)$ *is the information energy [16], which is defined as*

$$\mathcal{E}(P) = \sum_{i=1}^{n} p_i^2. \tag{21}$$

If we follow the same idea as for probabilistic divergence measures developed in [11] we obtain the following definition:

Definition 6. *Consider* f *a measure of fuzziness. Then, given a probability distribution* P *over* Ω *such that* $P(x_i) > 0$, *we define the* **probabilistic measure of fuzziness** *associated to* P *as*

$$f_P^1(A) = \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} (p_{(i)} - p_{(i-1)}) f(A_{(i)}), \tag{22}$$

where $\mathcal{E}(P)$ *is the Onicescu's informational energy and* $A_{(i)}$ *is defined by* $A_{(i)}(x_j) = min\{A(x_j), \Omega_i(x_j)\}$ *with*

$$\Omega_i(x_j) = \begin{cases} 1 & if \quad p(x_j) \geq p_{(i)} \\ 0 & si \quad p(x_j) < p_{(i)} \end{cases}$$

Proposition 5. f_P^1 *is a measure of fuzziness.*

Proof: Let us see that the three axioms of measure of Definition 4 hold:

- If A is a crisp subset, then $f_P(A) = \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} (p_{(i)} - p_{(i-1)}) f(A_{(i)}) = 0$. Note that we need that $P(x_i) > 0$ to have the reverse.
- For any fuzzy subset A, we have $M_{(i)} \succeq A_{(i)}$ and thus $f_P(M) \geq f_P(A)$.
- If $A \succeq B$, then $A_{(i)} \succeq B_{(i)}$ and thus $f_P(A) \geq f_P(B)$ ■

Another possibility for yhe definition of probabilistic measures of fuzziness is to use Theorem 4.

Definition 7. *Let us consider D a divergence measure in the conditions of Theorem 4 and a probability distribution P over Ω such that $P(x_i) > 0$. Then, we define the* **probabilistic measure of fuzziness** *associated to P as*

$$f_P^2(A) = D_P(Z, Z^c) - D_P(A, A^c). \tag{23}$$

Proposition 6. *f_P^2 is a measure of fuzziness.*

Proof: Trivial using Theorem 4 ■

It can be seen that these two definitions are not the same:

Example 6. *Consider $\Omega = \{x_1, x_2\}, c(x) = 1 - x$, and*

$$D(A,B) = \begin{cases} 1 \quad \text{if} \quad \text{A} \quad \text{crisp}, \quad \text{B} = \text{A}^\text{c} \\ 0 \qquad\qquad \text{if} \quad \text{A} = \text{B} \\ 0.5 \qquad\qquad \text{otherwise} \end{cases}$$

Finally, consider $P(x_1) = 0.8, P(x_2) = 0.2$.

Then, simple calculus lead us to

$$f_P^1 = \frac{1}{\mathcal{E}(P)} 0.1, f_P^2 = \frac{1}{\mathcal{E}(P)} 0.4.$$

However, we have the following:

Proposition 7. *Under the conditions of Theorem 4, if D is a local divergence measure, then $f_P^1 = f_P^2$.*

Proof:

$$f'_P(A) = \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} (p_{(i)} - p_{(i-1)}))(D(Z_{(i)}, Z_{(i)}^c) - D(A_{(i)}, A_{(i)}^c))$$

$$= \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} (p_{(i)} - p_{(i-1)})) \sum_{x_j / p(x_j) \geq p_{(i)}} (h(1,0) - h(A(x_j), A^c(x_j))$$

$= \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} (p_{(i)} - p_{(i-1)}))f(A_{(i)}) = f_P(A)$ ■

4 Conclusions

We feel that these results are interesting in order to translate properties from one of these families to another one. In Sections 2 and 3 we have seen an application of these relations. More examples can be found; this is part of our future research. More exactly, we feel that the relations between divergence measures and aggregation operations can be used in order to classify the first ones. It could also be interesting to study more deeply the probabilistic measures of fuzziness (see [12]). In [10], it can be found similar relations between aggregation operations and measures of fuzziness.

Acknowledgements

Research supported in part by DGES grants number PB97-1286 and PB97-1282. We would like also to thank an anonymous referee for very interesting remarks.

References

[1] P. Benvenutti, D. Vivona & M. Divari (1998). "Fuzziness Measures Via Pseudometrics on the Fuzzy Sets Space", *Proc. IPMU'98, Paris* (1998) 1036-1039.

[2] B. Bouchon-Meunier, M. Rifqi & S. Bothorel. "Towards general measures of comparison of objects", *Fuzzy Sets and Systems* **84** (1996) 143-153.

[3] B. Burbea & C.R. Rao. "Entropy differential metric, distance and divergence measures in probabilistic spaces: A unified approach", *T. Multiv. Anal.* **12** (1982) 575-596.

[4] D. Dumitrescu & H.F. Pop. "Degenerate and non-degenerate convex decomposition of finite fuzzy partitions-I", *Fuzzy Sets and Systems* **73** (1993) 365-376.

[5] D. Dubois & H. Prade. "Fuzzy Sets and Systems. Theory and Applications". *Academic Press.* (1980)

[6] J. N. Kapur. "Measures of informations and their applications". *John Willey and Sons.*(1994)

[7] A. Kaufmann. "Introduction to the Theory of Fuzzy Subsets". *Academic Press, New York.*(1975)

[8] G. Klir & T. Folger. "Fuzzy Sets, Uncertainty and Information". *Prentice Hall P T R.*(1988)

[9] A. De Luca & E. Termini. "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory", *Information and Control,* **20** (1972) 301-312.

[10] P. Miranda & P. Gil. "Una caracterización de medidas de borrosidad", *Proc. ESTYLF'98. Pamplona* (1998) 157-162. In Spanish.

[11] P. Miranda, E. Torres & P. Gil. "Probabilistic divergence measures", *Proc. EUROFUSE-SIC'99. Budapest.* (1999) 354-358.

[12] P. Miranda, E. Torres, S. Montes & P. Gil. "Medidas de borrosidad probabilísticas", *Proc. XXV Congreso Nacional de Estadística e Investigación Operativa. Vigo.* (2000). In Spanish.

[13] S. Montes. "Particiones y medidas de divergencia en modelos difusos", *Ph. Thesis. University of Oviedo.* (1998). In Spanish.

[14] S. Montes, I. Couso& C. Bertoluzza. "Fuzziness Measures from Divergences", *Proc. EUROFUSE-SIC'99, Budapest* (1999) 365-370.

[15] S. Montes, P. Gil & C. Bertoluzza. "Divergence measure between Fuzzy sets and fuzziness", *Proc. IPMU'98, Paris* (1998) 1030-1035.

[16] O. Onicescu. "Energie Informationelle", *C.R. Acad. Sci. Paris Ser. A,* **263** (1966) 841-842.

[17] L.A. Zadeh. "Fuzzy sets", *Information and Control,* **8** (1965) 338-353.