Divergence Measures and Aggregation Operations

Miranda, P.
Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo-Spain
est010@pinon.ccu.uniovi.es

Torres, E.
Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo-Spain
torres@etsiig.uniovi.es

Gil, P.
Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo-Spain
pedro@pinon.ccu.uniovi.es

Summary

In this paper we deal with the relations between divergence measures between two fuzzy sets and aggregation operations while having a finite universal set. We will prove that aggregation operations can be related with divergence measures. We feel that this can be useful in order to study some properties of these last measures. Finally, we find some relations between divergence measures and measures of fuzziness.

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

1 Introduction and basic concepts

In the following, divergence measures are denoted as D; the universal set is denoted by Ω; fuzzy subsets are denoted by A, B and so on; n denotes the cardinality of Ω; elements of Ω are denoted by x_i. An aggregation operation is denoted by s and a measure of fuzziness is denoted by f.

Although divergence measures and measures of fuzziness can attain the infinite value we will not consider this case (see remark 2 above).

Definition 1 *Let Ω be the universal set and let $\widetilde{\mathcal{P}}(\Omega)$ be the set of all 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** *[7] 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.

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 *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 $h : [0,1] \times [0,1] \mapsto \mathbb{R}$ verifying*

$$D(A, B) - D(A \cup \Omega^i, B \cup \Omega^i) = h(A(x_i), B(x_i)),$$

$$\forall A, B \in \widetilde{\mathcal{P}}(\Omega), \forall x_i \in \Omega \qquad (1)$$

where

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

For these special divergence measures the following can be proved [7]:

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

$$D(A, B) = \sum_{i=1}^{n} h(A(x_i), B(x_i)), \qquad (2)$$

verifying 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].*

The aggregation operations on fuzzy sets are operations by which several fuzzy sets are combined to produce a single set. In particular, fuzzy unions and fuzzy intersections are special cases of aggregation operations. There are a lot 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 [5].

Definition 3 *A* **aggregation operation** *is defined by a function* $s : [0,1]^n \mapsto [0,1]$ *verifying*

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

2 Relationship between divergence measures and aggregation operations

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

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

Let s *be an aggregation operation. Then,*

$$D(A,B) = ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n))) \quad (3)$$

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

Proof. Let s be an aggregation operation and let us see that D is a divergence measure.

- $D(A,B) = D(B,A)$ is trivial using condition 2 over h_i.
- $D(A,A) = ks(\vec{0}) = 0$ using condition 1 over h_i.
- $D(A \cup C, B \cup C) = ks(h_1((A \cup C)(x_1), (B \cup C)(x_1)), ..., h_n((A \cup C)(x_n), (B \cup C)(x_n)))$, $D(A,B) = ks(h_1(A(x_1), B(x_1), ..., h_n(A(x_n), B(x_n)))$. Let us see that $h_i((A \cup C)(x_i), (B \cup C)(x_i)) \leq h_i(A(x_i), B(x_i))$. We have three cases:
  1. If $C(x_i) \geq A(x_i)$, $C(x_i) \geq B(x_i)$, then $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(C(x_i), C(x_i)) = 0 \leq h_i(A(x_i), B(x_i))$.
  2. 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).
  3. 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 is a divergence measure ■

Let us give some remarks:

1. Condition 3 over h_i is needed in order to s can be applied. We could have considered $h_i(x,y) \leq 1$.
2. Divergence measures D obtained in the theorem verify that: If $h_i(A(x_i), B(x_i)) = h_i(C(x_i), D(x_i)), \forall i$, then $D(A,B) = D(C,D)$ as s is applied over the same vectors.
3. k is needed because divergence measures can reach values greater than 1.

Note that the conditions over h_i are the same as those for local divergence measures. Thus, we have the following

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

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

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

$$s(\vec{p}) = \frac{D(A,\emptyset)}{D(\Omega,\emptyset)} \quad (4)$$

with $A(x_i) = p_i$ *is an aggregation operation.*

Proof. Let us see that the axioms for aggregation operation hold:

- $s(\vec{1}) = \frac{D(\Omega,\emptyset)}{D(\Omega,\emptyset)}$ and thus $s(\vec{1}) = 1$. Now, $s(\vec{0}) = \frac{D(\emptyset,\emptyset)}{D(\Omega,\emptyset)}$ and then $s(\vec{0}) = 0$.
- Let $\vec{p} \leq \vec{q}$. 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

$$D(B,\emptyset) \geq D(B \cap A, \emptyset \cap A) = D(A,\emptyset).$$

Then s is an aggregation operation ■

Remark 1 *Maybe it is not possible to find s given D and h_i. To see this, consider $|\Omega| = 1, D(A,B) = |A(x) - B(x)|$ and*

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

Then, D can take infinite 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 $D(A,B) \neq D(C,D)$.

Theorem 1 and a generalisation of theorem 2 can be put together as:

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

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

and let D be a set function verifying that if $h_i(A(x_i), B(x_i)) = h_i(C(x_i), D(x_i)), \forall i$, then $D(A,B) = D(C,D)$. Then, D is a divergence measure if and only if it can be put as

$$D(A,B) = ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n))) \quad (5)$$

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

Proof.$\Rightarrow$ Let D and h_i be as in the theorem. Let us see that we can define s an aggregation operation. We define $s(\vec{p}) = \frac{D(A,\emptyset)}{D(\Omega,\emptyset)}$ with $A(x_i) = sup_{h_i(0,x') \leq p_i} x'$.

- $s(\vec{1}) = \frac{D(\Omega,\emptyset)}{D(\Omega,\emptyset)}$ and thus $s(\vec{1}) = 1$. Now, $s(\vec{0}) = \frac{D(\emptyset,\emptyset)}{D(\Omega,\emptyset)}$ and then $s(\vec{0}) = 0$.
- Let $\vec{p} \leq \vec{q}$. Then, if we define $A(x_i) = sup_{h_i(0,x') \leq p_i} x', B(x_i) = sup_{h_i(0,x') \leq q_i} x'$, 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

$$D(B,\emptyset) \geq D(B \cap A, \emptyset \cap A) = D(A,\emptyset).$$

$\Leftarrow$ Theorem 1 ■

Remark 2 *Divergence measures can assume the ∞ value. Consider, for example $D(A,B) = \sum_{x \in \Omega} |A(x) - B(X)|$ when $|\Omega| = \infty$. However, in the finite case it is not really a useful measure.*

3 Relationship between divergence measures and measures of fuzziness

Measures of fuzziness appear in the seventies to measure the vagueness or fuzziness of a fuzzy set. They have been extensively studied [2], [10].

Definition 4 *A* **measure of fuzziness***[6] is a mapping*

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

verifying

- *$f(A) = 0$ if and only if A is a crisp subset.*
- *If $A \preceq B$ then $f(A) \leq f(B)$ where $A \preceq B$ denotes 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.*

We will consider $A \preceq B \Leftrightarrow |A(x) - A^c(x)| \leq |B(x) - B^c(x)|, \forall x \in \Omega$ with c a fuzzy complement.

Now, the following can be proved:

Theorem 4 *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*

$$f(A) = k(D(Z,Z^c) - D(A,A^c)), k > 0 \quad (6)$$

with Z a crisp subset and D a normalized divergence measure verifying

- *$D(A,A^c) = 1 \Leftrightarrow A$ 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 $D(A,B) = D(C,D)$.*

Proof.$\Rightarrow$Let f be a measure of fuzziness. Let us see that given $A,B, \exists C | D(C,C^c) = D(A,B)$. 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 $x \in \Omega$ there exists v such that $c(v) - v = |A(x) - B(x)|$. This value of v is unique: Suppose that there exists another value $v' > v$ verifying $c(v') - v' = |A(x) - B(x)|$. Then, as c is a non-increasing function, $c(v') \leq c(v)$ and thus $c(v') - v' < c(v) - v$ contradicting our hypothesis. We define $C_{A,B}(x) = v$. Using our hypothesis 3 over D we obtain that $D(C_{A,B}, C_{A,B}^c) = D(A, B)$.

We define $D(C_{A,B}, C_{A,B}^c) = \frac{f(M) - f(C_{A,B})}{k}$. By hypothesis 3, D is well-defined. Now, let us see that D is a divergence measure.

- 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 $D(A, A) = 0$.
- $D(A, B) = D(B, A)$ is trivial.
- It is trivial. 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 thus $D(A, B) \geq D(A \cup D, B \cup D)$. The same holds for the intersection.

Then D is a divergence measure. It is easy to see that it verifies the conditions of the theorem.

Note also that for a crisp subset Z, given Z, Z^c, $C_{Z,Z^c} = \emptyset$ and then $D(Z, Z^c) = f(M)$. Thus, $f(A) = D(Z, Z^c) - D(A, A^c)$.

$\Leftarrow$Let D be as in the theorem and let us see that f is a measure of fuzziness.

- If A crisp, $f(A) = 0$ by hypothesis 1 and these are the only subsets verifying this condition.
- Let $A \preceq B$, then $D(A, A^c) = D(A', A'^c)$ with

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

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

Let us see some remarks:

1. If we consider general not-normalized divergence measures k can be removed.
2. We could make the same process as in the first part of the paper with general operators h_i instead of the absolute value. Then, we would have obtained similar results.
3. Note that all hypothesis are verified by the usual measures of fuzziness and divergence measures. In fact, $f(A) = f(A^c)$ is considered as an axiom by many authors [3].
4. This result was proved by Montes *et al.* in [8] for local divergence measures. Our result is an extension of this one: To prove that we have really an extension we have to find a non-local divergence measure verifying the conditions of the theorem. It is easy to see that for $\Omega = \{x_1, x_2\}$,

   $$D(A, B) = (1 + |A(x_1) - B(x_1)|)(1 + |A(x_2) - B(x_2)|) - 1$$

   is a divergence measure in these conditions.

Acknowledgements

This research has been supported in part by DGES grant number PB97-1286.

References

[1] B. De Baets & R. Mesiar (1998). T-partitions. *Fuzzy Sets and Systems 97* pp 213-223

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

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

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

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

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

[7] S. Montes (1998) Particiones y medidas de divergencia en modelos difusos.*Ph. Thesis. University of Oviedo.*

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

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

[10] E. Trillas & T. Riera. Entropies in finite fuzzy sets. *Inf. Sci. 15* pp 159-168.