
Divergence measures: from uncertainty to imprecision

Susana Montes¹, Susana Díaz¹, and Davide Martinetti²

¹ Dept. of Statistics and Operational Research, University of Oviedo, Spain
`{montes,diazsusana}@uniovi.es`

² INRA-PACA, UR Ecodeveloppement, France `davide.martinetti@inra.fr`

Summary. The link between information theory and fuzzy logic has been proven in several previous papers. From this starting point, we propose here a review about the concept of divergence measures, which was proposed as a tool for comparing two fuzzy sets. The initial definition comes from the ideas behind the classical concept of divergence between two probability distributions. Following a path similar to the one considered to obtain fuzziness measures from uncertainty measures, we are able to define fuzzy divergences. Apart from that, some possible generalizations are considered.

Summary. *The link between information theory and fuzzy logic has been proven in several previous papers. From this starting point, we propose here a review about the concept of divergence measures, which was proposed as a tool for comparing two fuzzy sets. The initial definition comes from the ideas behind the classical concept of divergence between two probability distributions. Following a path similar to the one considered to obtain fuzziness measures from uncertainty measures, we are able to define the fuzzy divergences. Apart from that, some possible generalizations are considered.

1 Introduction

Dealing with lack of information is a usual problem in many areas. This lack of information can be given in two different ways: uncertainty or imprecision. In the first case, we deal with experiments where we can have more than one possible outcome, each possible outcome can be specified in advance, but the outcome of the experiment depends on chance. For instance, in a coin toss, we know the two possible outcomes, head or tail, but we do not know the final result. In the second case, we have no uncertainty about the result of the experiment but imprecision. Thus, for instance, if we consider again the experiment of the coin toss, the coin could be already thrown but maybe it is too old and we are not sure that the face it shows is clearly a head.

Information theory studies the quantification and communication of the information and, in particular, it measures the amount of uncertainty involved in the value of a random experiment. It was originally proposed by Shannon in 1948 ([30]) as a tool in signal processing. Thus, this theory combines a lot of different fields such as mathematics, statistics, computer science, physics and electrical engineering. From the beginning, this theory was revealed as an interesting tool in many other areas and therefore a lot of researchers started to work on it (Rényi [29], Oniçescu [26], Sharma and Mittal [31], Havrda and Charvát [11], etc.). Later, an important step was given by Kampé de Fèriet and Forte [12] with an axiomatic definition of the information with or without a probability measure. From the theoretical aspects of this theory, Kullback [17] found a lot of interesting applications in statistical inference. From this initial application a lot of papers have been developed in this area. In particular, some very important achievements have been obtained by Pardo (see, among others, [27, 28]). An important review about all these theories can be found in Gil [10], since he was one of the most important researchers in this area in Spain. Divergence measures between probability distribution were an important topic on this monograph and it is the starting point of this chapter, as we will see later.

On the other hand, Zadeh [33] introduced in 1965 the concept of fuzzy set, as a way to model vague or poorly defined properties for situations in which it is not possible to fully discriminate between having and not having said properties. From that, a whole mathematical and applied theory to deal with imprecision was developed. It is known as Fuzzy Logic Theory. Two interesting monographs about this theory were written by Dubois and Prade [6] and Klir and Folger [13].

As we can see from the title of this last book, the concepts fuzzy sets, uncertainty and information are mixed. This is not by chance, since these topics are very related, as we can see in [7, 8, 9]. In particular, we have studied ([25]) the relationship between uncertainty measures defined in Information Theory ([12]) and the fuzziness measures introduced by De Luca and Termini ([5]) and later analyzed in a deeper way by Knopfmacher ([14]). The link between measures of uncertainty and imprecision in fuzzy environments will lie in what we will refer as divergence measure, because of the analogy with the classical meaning of the term used in comparing two probability distributions (see, for instance, [28]). The main purpose of this chapter is to use these measures to compare two fuzzy sets.

As introductory notions, we present two axiomatic definitions to measure the entropy – uncertainty measures and fuzziness measures in Section 2. A study on the relationship between them, in the most general context, is given there. The definition of divergence measure between fuzzy sets is given in Section 3 following the ideas considered previously. The most important results are contained in that section where we also comment some extensions. Finally, we conclude the work with some comments in Section 4.

2 Preliminary

Necessary concepts to understand the remaining parts of this work are given in this section. In particular, we will focus on the definitions and notations for uncertainty measures and fuzziness measures.

2.1 Uncertainty measures

The first probabilistic uncertainty measure (also called entropy) was given by Shannon ([30]) in the context of Communication Theory. That initial definition considered that the uncertainty for a random experiment can be measured by means of the quantity

$$H(P) = - \sum_{i=1}^n p_i \log_2(p_i)$$

where the values p_i represent the probabilities of the possible results of the experiment.

From that initial definition, a lot of generalization have been proposed in the literature. Thus, Salicru et al. (see [27]) proved that all these measures of entropy are part of a wider family, which are named $h - \phi$ -entropies. This family is slightly more general than Ben Bassat's family of f -entropies that were defined as those functions that can be expressed like

$$H(P) = \sum_{i=1}^n f(p_i)$$

where f is a concave function. Later, the quasi- ϕ -entropies were introduced and characterized in the case of discrete distributions ([2]). Thus, it is a family more general than Ben Bassat's one but different from the family of $h - \phi$ -entropies. More precisely, they are defined by

$$H(P) = \sum_{i=1}^n \phi(p_i)$$

where ϕ is a function such that $\phi(\lambda x + (1 - \lambda)y) \geq \lambda\phi(x) + (1 - \lambda)\phi(y), \forall x, y \in [0, 1], x + y \leq 1$.

An important property of uncertainty measures is the Principle of Transfer or Pigou-Dalton's condition. An uncertainty measure H fulfils this property if given two probability distributions P and P' with parameters (p_1, p_2, \dots, p_n) and $(p'_1, p'_2, \dots, p'_n)$ respectively, then $H(P) \leq H(P')$, where, $p_k = p'_k, \forall k \notin \{i, j\}$ and $p'_i = p_i + \delta, p'_j = p_j - \delta$ for some $\delta \leq (p_i - p_j)/2$. It is a very logical property, since it means that the more similar the probabilities of two outcomes of an experiment are, the higher the uncertainty is.

2.2 Fuzziness measures

After having commented some results about uncertainty measures or probabilistic entropies, we now introduce fuzzy sets and the measures of their fuzziness, i.e. the non-probabilistic entropies. They are well-known and can be found in a wide range of sources (see, for instance, the classical books [6, 13]).

The universal set is denoted by X . A fuzzy subset of X is a mapping from X into the unit interval $[0, 1]$. In this framework, we use the following notations:

- $\mathcal{P}(X)$ is the set of all subsets of X ,
- $\mathcal{F}(X)$ is the set of all fuzzy subsets of X ,
- $A \in \mathcal{P}(X)$ will denote any crisp set,

- $\tilde{A} \in \mathcal{F}(X)$ will denote any fuzzy set.

We identify a fuzzy set and its membership function. Thus we have that $X(x) = 1$ for all $x \in X$ and for the empty set we have $\emptyset(x) = 0$ for all $x \in X$.

Two further important concepts are the containment relation and the complement set. We consider the standard Zadeh's negation for the complement (see [33]).

Definition 1. Let $\tilde{A}, \tilde{B} \in \mathcal{F}(X)$. The complement of \tilde{A} is the fuzzy set $\tilde{A}^c(x) = 1 - \tilde{A}(x)$, $x \in X$. \tilde{A} is contained in \tilde{B} , denoted by $\tilde{A} \subseteq \tilde{B}$ if $\tilde{A}(x) \leq \tilde{B}(x)$ for all $x \in X$.

Apart from the previous relation of containment, we consider the concepts of intersection and union of fuzzy sets. The initial definitions were also given in [33] by means of the minimum and the maximum operators. However, they are not the only way to generalize the classical set operations, since there exist a broader class of functions to represent them. In fact, for the intersection, this class is referred as t-norm and for the union as t-conorm.

A triangular norm (t-norm) is a function $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ satisfying the following properties:

- (T1) $T(a, b) = T(b, a)$, for all $a, b \in [0, 1]$,
- (T2) $T(T(a, b), c) = T(a, T(b, c))$, for all $a, b, c \in [0, 1]$,
- (T3) $b \leq c \Rightarrow T(a, b) \leq T(a, c)$, for all $a, b, c \in [0, 1]$,
- (T4) $T(a, 1) = a$, for all $a \in [0, 1]$.

Some important examples of t-norms are:

- Minimum: $T_M(a, b) = \min(a, b)$, for all $a, b \in [0, 1]$,
- Product: $T_P(a, b) = a \cdot b$, for all $a, b \in [0, 1]$,
- Łukasiewicz t-norm: $T_L(a, b) = \max(a + b - 1, 0)$, for all $a, b \in [0, 1]$,
- Drastic t-norm:

$$T_D(a, b) = \begin{cases} \min(a, b), & \text{if } \max(a, b) = 1 \\ 0, & \text{otherwise.} \end{cases}$$

For these basic t-norms, it holds that $T_D \leq T_L \leq T_P \leq T_M$. In fact, for any t-norm T , it is true that $T_D \leq T \leq T_M$. By changing the neutral element from 1 to 0, we obtain the triangular conorms (t-conorm). A t-norm T and a t-conorm S are dual iff for each $a, b \in [0, 1]$ it holds that $T(a, b) = 1 - S(1 - a, 1 - b)$.

The duals conorms of the t-norms presented earlier are the following:

- Maximum: $S_M(a, b) = \max(a, b)$, for all $a, b \in [0, 1]$,
- Probabilistic sum: $S_P(a, b) = a + b - a \cdot b$, for all $a, b \in [0, 1]$,
- Łukasiewicz t-conorm: $S_L(a, b) = \min(a + b, 1)$, for all $a, b \in [0, 1]$,
- Drastic t-conorm:

$$S_D(a, b) = \begin{cases} \max(a, b), & \text{if } \min(a, b) = 0. \\ 1, & \text{otherwise.} \end{cases}$$

Using t-norms and t-conorms, we can define the intersection and union of two fuzzy as follows.

Definition 2. Let $\tilde{A}, \tilde{B} \in \mathcal{F}(X)$. Given a t-norm T and a t-conorm S ,

- $\tilde{A} \cap \tilde{B}(x) = T(\tilde{A}(x), \tilde{B}(x)), \forall x \in X;$
- $\tilde{A} \cup \tilde{B}(x) = S(\tilde{A}(x), \tilde{B}(x)), \forall x \in X.$

Thus, we can denote by (X, T, S) the triple formed by the universe with the t-norm and the t-conorm defining the intersection and the union, respectively.

The entropy for a fuzzy sets is quantified by means of the non-probabilistic entropies or fuzziness measures (see, for instance, [32]), which are defined as follows.

Definition 3. *A fuzziness measure is a real function f defined on $\mathcal{F}(X)$, fulfilling the following requirements:*

1. $f(\tilde{A}) = 0 \iff \tilde{A}$ is a crisp set.
2. If $\tilde{A}, \tilde{B} \in \mathcal{F}(X)$ and \tilde{A} is “sharper” than \tilde{B} , then $f(\tilde{A}) \leq f(\tilde{B})$.
3. $f(\tilde{A})$ takes maximum value if and only if \tilde{A} is “maximally fuzzy”.

This last definition is based on the concepts “sharper than” and “maximally fuzzy”, although the second one follows from the former. Thus, the most usual criteria to define the relation “to be sharper than” are the following:

- \tilde{A} is sharper than \tilde{B} iff either $\tilde{A}(x) \leq \tilde{B}(x) \leq 1/2$ or $\tilde{A}(x) \geq \tilde{B}(x) \geq 1/2$ for any x in X (see [13]) or
- \tilde{A} is sharper than \tilde{B} iff $|\tilde{A}(x) - 1/2| \leq |\tilde{B}(x) - 1/2|$ for any x in X (see [6]).

It is clear that the first one is a particular case of the second one and therefore we are going to consider the most general definition.

Knopfmacher introduced in 1975 a very important family of fuzziness measure, the Knopfmacher’s class ([14]), which is given by the functions f such that

$$f(\tilde{A}) = F \left(\sum_{x \in X} c_x \cdot g_x(\tilde{A}(x)) \right)$$

for any \tilde{A} in $\mathcal{P}(X)$ where $c_x \in \mathbb{R}^+$; g_x is a real-valued function such that $g_x(0) = g_x(1) = 0, g_x(t) = g_x(1-t), \forall t \in [0, 1]$ and g_x is strictly increasing on $[0, 1/2]$; F is a positive strictly increasing function with $F(0) = 0$.

Later, we consider a particular class of Knopfmacher fuzziness measure (see [21, 23]) when F is the identity, g_x is the same for all $x \in X$ (we denoted g_x by u_f or simply u) and u is concave. Any function in this family was named local fuzziness measure.

2.3 From uncertainty to fuzziness

Proposition 1. ([25]) *Let (X, \mathcal{A}, μ) a measurable space and let H be an uncertainty measure fulfilling the Pigou-Dalton’s condition and such that $H(P) = 0 \iff P$ is degenerate. The map f defined as follows*

$$\begin{aligned} f : \mathcal{A}^* &\longrightarrow \mathbb{R}^+ \\ \tilde{A} &\longrightarrow \int_X H(\tilde{A}(x), \tilde{A}^c(x)) d\mu(x) \end{aligned}$$

is a fuzziness measure and it belongs to the Knopfmacher’s class.

If we work on some particular spaces, we are also able to establish a one-to-one correspondence between fuzziness measures and uncertainty measures. Thus, if we consider the subset of uncertainty measures given by

$$\mathcal{H}_2 = \{H \mid H \text{ is a quasi-}\phi\text{-entropy with } \phi \text{ continue, } \phi(x) = \phi(1-x), \\ \forall x \in [0, \frac{1}{2}] \text{ and } \phi(x) = 0 \Leftrightarrow x = 0\}.$$

, we have the injective property, as we can see in the following proposition.

Proposition 2. ([25]) *If F_1 is a map from \mathcal{H}_2 in \mathcal{F} such that*

$$F_1(H)(\tilde{A}) = \int_X H(\tilde{A}(x), \tilde{A}^c(x)) d\mu(x)$$

where \mathcal{F} denotes the Knopfmacher's class of fuzziness measures, then we have that F_1 is injective.

If we restrict our study to the family of ϕ -entropies given by $\mathcal{H}_\phi = \{H \in \mathcal{H}_2 \mid \phi \text{ is concave}\}$ and the family of fuzziness measures given by $\mathcal{F}_1 = \{f \in \mathcal{F} \text{ with } g \text{ continue}\}$ we have the bijection.

Theorem 1. ([25]) *There exists an one-to-one correspondence between the family of uncertainty measures \mathcal{H}_ϕ and the family of fuzziness measures \mathcal{F}_1 .*

3 Divergence measures

From the previous section, we could notice that the imprecision about the membership of any element $x \in X$ in a fuzzy set \tilde{A} could be represented by a probability distribution $\{\tilde{A}(x), \tilde{A}^c(x)\}$. Then, we looked at the classical divergence measures between probability distributions (see, for instance, [10, 28]) to try to compare two fuzzy sets. Thus, from this starting point, we proposed a new way to compare two fuzzy set ([21]), the divergence, with the following properties:

- It becomes zero when the two sets coincide.
- It is a nonnegative and symmetric function.
- It decreases when the two subsets become “more similar” in some sense.

While it is easy to formulate the first and the second conditions analytically, the third one depends on the formalization of the concept “more similar”. We base our approach on the fact that if we add a subset \tilde{C} to both fuzzy subsets \tilde{A}, \tilde{B} , we obtain two subsets which are closer to each other; the same with the intersection.

Definition 4. *Let (X, T, S) be a triple with X a universe and T and S any t -norm and t -conorm, respectively. A map $D : \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow \mathbb{R}$ is a divergence measure with respect to (X, T, S) iff for all $\tilde{A}, \tilde{B} \in \mathcal{F}(X)$, D satisfies the following conditions:*

1. $D(\tilde{A}, \tilde{A}) = 0$;
2. $D(\tilde{A}, \tilde{B}) = D(\tilde{B}, \tilde{A})$;
3. $\max\{D(\tilde{A} \cup \tilde{C}, \tilde{B} \cup \tilde{C}), D(\tilde{A} \cap \tilde{C}, \tilde{B} \cap \tilde{C})\} \leq D(\tilde{A}, \tilde{B})$, for all $\tilde{C} \in \mathcal{F}(X)$, where the union and intersection are defined by means of S and T , respectively.

It is clear that a divergence measure is associated to a triple (X, T, S) and a map D can be a divergence measure with respect to a t-norm and it can be not a divergence measure with respect to a different t-norm. However, when there is not ambiguity, we will say just a divergence measure without specifying the used t-norm and t-conorm.

After different studies of this concept [3, 21, 23, 24, 25], we presented the most general study in [15], where we can also find the following examples.

Example 1. ([15]) The map

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

is a divergence for any triple (X, T, S) .

On the other hand, if we consider the map

$$D(\tilde{A}, \tilde{B}) = \sum_{x \in X} \alpha_x \cdot |\tilde{A}(x) - \tilde{B}(x)|$$

where $\alpha_x \geq 0$ for any $x \in X$, $\sum_{x \in X} \alpha_x = 1$ and X is a finite space, D is a divergence for the minimum t-norm, the product t-norm or the Lukasiewicz t-norm, but it is not for the drastic t-norm.

A divergence measure can be seen as a particular case of dissimilarity when the minimum t-norm is considered, which is the most usual way to compare two fuzzy sets [18]. Moreover, it avoids some counterintuitive examples for dissimilarities, while both divergence and dissimilarity measures can be seen as a particular case of the general measures of comparison given by Bouchon-Meunier in 1996 ([1]). An interesting study about different ways to compare fuzzy sets can be found in [4].

From this starting point, we have been able to generalize this concept to define the divergence measure for comparing two intuitionistic fuzzy sets [19]. The particular case of local divergences for intuitionistic fuzzy sets was studied in [20]. There, we presented interesting applications of this concept in Pattern Recognition and Decision Theory. A similar generalization have been done for hesitant fuzzy sets in [16]. Moreover, we have been able to use the divergences to measure the fuzziness of a fuzzy set by comparing it with the closest crisp set and conversely, we have used fuzziness measures to define a divergence measure [22]. All these definitions and results can be considered as a heritage of the classical divergence measures, and more precisely, of the knowledge about them conveyed by Prof. Gil to the authors of this work.

4 Conclusion

In this paper we have studied some relationships among different ways to compare two elements, under uncertainty and imprecision. Thus, we have used the classical divergence measures between two probability distributions to obtain a new way to compare two fuzzy sets. This is a particularly interesting case of dissimilarity in some cases and it has very interesting and specific properties. The link between randomness and fuzziness is proven one more time, as we did previously for probabilistic and non-probabilistic entropies.

Acknowledgement. We would like to acknowledge the help and support of Prof. Gil to initiate us to the wonderful world of research. He used to say he was our scientific father and we are honored he really was.

From the economical point of view, this work was partially supported by the research project TIN2014-59543-P (Spain).

References

1. Bouchon-Meunier B, Rifqi M, Bothorel S (1996) Towards general measures of comparison of objects, *Fuzzy Sets and Systems* 84 , 143–153
2. Couso I, Gil P (1998). Characterization of a family of entropy measures. In: *Proceedings of IPMU'98 Conference*, pp. 1030–1035
3. Couso I, Montes S (2008) An axiomatic definition of fuzzy divergence measures, *International Journal of Uncertainty, Fuzziness and Knowledge-based systems* 16/1, 1–17
4. Couso I, Garrido L, Sánchez L (2013) Similarity and dissimilarity measures between fuzzy sets: A formal relational study, *Information Sciences* 229, 122–141
5. De Luca A, Termini S (1972) A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory, *Information and Control* 20: 301–312
6. Dubois D, Prade H (1980) *Fuzzy Sets and Systems: Theory and Applications*. Academic Press
7. Gil MA, Gil P (2015) Randomness and Fuzziness: Combined Better than Unified. In: Magdalena, Verdegay, Esteve (eds) *Studies in Fuzziness and Soft Computing Volume 322*. Springer
8. Gil MA, Lopez M, Gil P (1985) Quantity of information; comparison between information systems: 1. Non-fuzzy states, *Fuzzy Sets and Systems* 15: 65–78
9. Gil MA, Lopez M, Gil P (1985) Quantity of information; comparison between information systems: 2. Fuzzy states, *Fuzzy Sets and Systems* 15: 129–145
10. Gil P (1981) *Teoría matemática de la información*. I.C.E.
11. Havrda J, Charvát F (1967) Quantification method of classification processes. Concept of structural α -entropy, *Kybernetika* 3-1, 30–35
12. Kampé de Fériet J, Forte B (1967) Information et probabilité, *C. R. Acad. Sci. Paris* 265: 110-114, 142–146, 350–353
13. Klir GJ, Folger TA (1988) *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall
14. Knopfmacher J (1975) On Measures of Fuzziness, *J. Math. Analysis and Applications* 49: 529–534
15. Kobza V, Janis V, Montes S (2017) Generalized local divergence measures, *Journal of Intelligent and Fuzzy Systems* 33, 337–350
16. Kobza V, Janis V, Montes S (2017) Divergence measures on hesitant fuzzy sets, *Journal of Intelligent and Fuzzy Systems* 33, 1589–1601
17. Kullback S (1959) *Information Theory and Statistics*. John Wiley and Sons, NY
18. Lui X (1992) Entropy, distance measure and similarity measure of fuzzy sets and their relations, *Fuzzy Sets and Systems* 52, 305–318
19. Montes I, Pal N, Janis V, Montes S (2015) Divergence measures for intuitionistic fuzzy sets, *IEEE Transactions on Fuzzy Systems* 23/2, 444–456

20. Montes I, Pal N, Janis V, Montes S (2016) Local divergences for Atanassov intuitionistic fuzzy sets, *IEEE Transactions on Fuzzy Systems* 24/2, 360–373
21. Montes S, Gil O (1998) Some Classes of Divergence Measures Between Fuzzy Subsets and Between Fuzzy Partitions, *Mathware and Soft Computing* 5, 253–265
22. Montes S, Couso I, Bertoluzza C (1998) Some classes of fuzziness measures from local divergences, *Belgian Journal of Operations Research, Statistics and Computer Science* 38, 37–49
23. Montes S, Gil P, Bertoluzza C (1998) Divergence between fuzzy sets and fuzziness. In *Proceedings of IPMU'98 Conference*, pp. 1030–1035
24. Montes S, Couso I, Gil P, Bertoluzza C (2002) Divergence measures between fuzzy sets, *International Journal of Approximate Reasoning* 30/2, 91–105
25. Montes S, Couso I, Jimenez J, Gil P (2005) Las medidas de incertidumbre probabilística y no probabilística como herramienta en la comparación de conjuntos. In: *Volumen homenaje al profesor Ildefonso Yáñez de Diego, UNED*, pp. 59–74
26. Oniçescu O (1966) *Theorie de l'information. Energie informationelle*, C. R. Acad. Sci. Paris, Ser. A 263: 841–842
27. Pardo L (1997) *Teoría de la Información Estadística*. Hespérides
28. Pardo L (2006) *Statistical Inference Based on Divergence Measures*. Chapman & Hall
29. Rényi A (1966) *Calcul des Probabilités*. Dunod
30. Shannon CE (1948) A mathematical Theory of Communication, *Bell Syst. Tech. J.* 27: 379–423, 623–656
31. Sharma BD, Mittal DP (1975) New nonadditive measures of entropy for discrete probability distributions, *J. Math. Sci.* 10, 28–40.
32. Trillas E, Riera T (1978) Entropies in Finite Fuzzy Sets, *Information Sciences* 15: 159–168
33. Zadeh L (1965) Fuzzy Sets, *Inform. Control* 8: 338–353