Decision Making and Divergence Measures: An Application

Pedro Miranda, Emilio Torres and Pedro Gil

Department of Statistics and Operational Research
University of Oviedo, C/Calvo Sotelo s/n
33007 Oviedo, SPAIN
E-mail: est010@pinon.ccu.uniovi.es

Abstract

In this paper we deal with divergence measures between two actions in decision-making models when we have the information given by utility values. A definition of divergence in which we generalize fuzzy measures is proposed and we show the relation among divergence measures, fuzzy measures and utility. We also propose a new concept using the information of a probability distribution. Finally, the results are applied to a practical example.

Keywords: decision-making, divergence measure, fuzzy measure, information energy.

1 Introduction

In multicriteria decision making and pattern recognition, fuzzy approaches have been shown to be superior to traditional methods, particularly when the criteria and features are not well defined [7]. What makes fuzzy set theory and fuzzy logic so attractive is the fact that they provide a powerful and flexible framework for representing vague and ill-defined concepts.

Decision-making theory has developed important models for making a rational selection among alternative courses of action when information is incomplete and/or uncertain. However, the classical principles and methods for making the best decision fail when the two or more alternative courses of action are not clearly defined or well differenced.

To illustrate this, consider the next investment decision problem [8]: Tom has decided to invest \$1000 for a year, and a broker has selected four potential investments she believes would be appropriate for Tom: gold, a junk bond, a growth stock and a certificate deposit. Let us define the states of the world qualitatively as a large rise, a small rise, no change, a small fall and a large fall in the stock market

Actions	States of Nature				
	Large Rise	Small Rise	No Change	Small Fall	Large Fall
Gold	-100	100	200	300	0
Bond	250	200	150	-100	-150
Stock	500	250	100	-200	-600
C-D	10	10	10	10	10

Table 1: Payoff table

Decision Alternatives	States of Nature					Expected Utility
	Large Rise	Small Rise	No Change	Small Fall	Large Fall	
Gold	0.35	0.65	0.75	0.90	0.5	0.63
Bond	0.85	0.75	0.70	0.35	0.30	0.67
Stock	1.00	0.85	0.65	0.25	0	0.675
C-D	0.50	0.50	0.50	0.50	0.50	0.50
Probability	0.20	0.30	0.30	0.10	0.10	

Table 2: Decision utility payoff table

over the next year. The specific payoffs for the four different investments, based on the broker's analysis, are given in Table 1.

Now suppose that Tom's broker offers the following projections based on past stock market performance: $P(\text{LargeRise}) = 0.2$, $P(\text{SmallRise}) = 0.3$, $P(\text{NoChange}) = 0.3$, $P(\text{SmallFall}) = 0.1$, $P(\text{LargeFall}) = 0.1$. Let us consider utility values as membership functions of each action. These values are given in Table 2. In fact, actions are considered as fuzzy subsets.

The way of determining the utility values corresponding to payoffs has been based on the indifference probabilities [8]. It consists in assigning a utility 0 to the lowest value and a utility 1 to the highest one, and for all other possible payoff values the decision maker is asked as follows: "Suppose you could receive this payoff for sure, or alternatively, you could receive either the highest payoff with probability p and the lowest payoff with probability $(1 - p)$. What value of p would make you indifferent between these two situations?". The answers to these questions are called "indifference probabilities for the payoffs" and are used as the utility values. Obviously, large payoffs have large utility values, but the relative scale of utility values can differ vastly from that of the payoff values.

Since the decision with the highest expected utility is the stock investment, it would be selected using the expected utility criterion. Comparing the expected utilities of stock investment and gold, we note that there is little difference between the two values. In fact, with slightly different utility values, we could have chosen gold as the best decision.

In this paper, we propose a method to deal with the problem of decision-making under conditions of fuzzy actions, where the decision maker does not distinguish clearly among two or more actions. The basic objective of divergence measures is to reduce the indefiniteness concerning acts that are ill-differenced. Divergence measures [2] are used in order to quantify the difference between two fuzzy subsets of a universal set Ω. These measures try to maintain the properties of classical divergence measures between two probability distributions which appear in Information Theory [6, 12]. From this concept, it is possible to define divergence measures between fuzzy partitions [9, 3, 4]. We can also use divergence measures to generate measures of fuzziness following the study of difference between a fuzzy subset and its complementary [10].

In this paper, we also deal with divergence measures between two fuzzy sets when we have the information given by a probability measure. In this situation, it makes sense to use such information. A definition of a divergence measure in which we use probabilistic information is proposed for the finite case and f or the countable one. We start with the properties that this measure must verify; we call these new measures probabilistic divergences and we also study some of their properties. Probabilistic divergences can be considered as a correction of divergence measures. In fact, when a uniform distribution is considered, the original divergence measure and the probabilistic divergence will be the same function.

The paper is structured as follows: the first section is devoted to the basic concepts; then, the relation between fuzzy measures and divergence measures is studied. After this, the notion of probabilistic divergence is introduced. We finish with an example and conclusions.

2 Basic concepts

In the following, fuzzy subsets of a universe Ω are denoted by A, B and standard fuzzy union and intersection are considered, i.e. $(A \cup B)(x) = \max\{A(x), B(x)\}$ and $(A \cap B)(x) = \min\{A(x), B(x)\}$. $\widetilde{\mathcal{P}}(\Omega)$ denotes the set of all fuzzy subsets of Ω.

Definition 1 [9] *A* **divergence measure** *D (on Ω) is a $\widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \mapsto \overline{\mathbb{R}}$ mapping satisfying the following axioms, for any $A, B, C \in \widetilde{\mathcal{P}}(\Omega)$:*

d1. $D(A, B) = D(B, A)$;

d2. $D(A, A) = 0$;

d3. $\max\{D(A \cup C, B \cup C), D(A \cap C, B \cap C)\} \leq D(A, B)$.

The following properties can be shown easily:

1. $D(A, B) \geq 0$;
2. If $A \subseteq B \subseteq C \subseteq D$, then $D(A, B) \leq D(A, C)$ and $D(B, C) \leq D(A, D)$.

Definition 2 [9] *A divergence measure D is called* **local** *if there exists a mapping $h : [0,1] \times [0,1] \mapsto \mathbb{R}$ such that*

$$D(A,B) - D(A \cup \Omega^i, B \cup \Omega^i) = h(A(x_i), B(x_i)),\ \forall A, B \in \widetilde{\mathcal{P}}(\Omega),\ \forall x_i \in \Omega$$

where

$$\Omega^i(x) = \begin{cases} 1 & \text{if } x = x_i \\ 0 & \text{otherwise} \end{cases}.$$

In fact, for a local divergence measure each coordinate is independent of the others and all coordinates are all equally important. The following result holds:

Proposition 1 [9] *Let $\Omega = \{x_1, \ldots, x_n\}$. A divergence measure D is local if and only if there exists a mapping $h : [0,1] \times [0,1] \mapsto \mathbb{R}$ such that*

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

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 decreasing function on $[0,y]$ and increasing on $[y,1]$, $\forall y \in [0,1]$.*

Definition 3 *A divergence measure D is called* **normal** *if it satisfies axioms* **d1**–**d3** *and the following axioms:*

d4. $D(\Omega, \emptyset) = 1$;

d5. $D(A \cup B, A \cap B) \geq D(A,B), \forall A,\ B \in \widetilde{\mathcal{P}}(\Omega)$.

Definition 4 [13] *A* **fuzzy measure** *μ on $\mathcal{P}(\Omega)$ is a $\mathcal{P}(\Omega) \mapsto [0,1]$ mapping satisfying the following axioms, for any*

1. $\mu(\emptyset) = 0,\ \mu(\Omega) = 1$;
2. $A \subset B$ *implies* $\mu(A) \leq \mu(B)$.

μ is called **additive** *if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$.*

Well-known examples of additive fuzzy measures are classical probability measures. We can easily extend the above definition to fuzzy subsets as follows:

Definition 5 [1] *A* **fuzzy measure** *on $\widetilde{\mathcal{P}}(\Omega)$ is a mapping $\mu : \widetilde{\mathcal{P}}(\Omega) \mapsto [0,1]$ satisfying the following axioms, for any $A, B \in \widetilde{\mathcal{P}}(\Omega)$:*

1. $\mu(\emptyset) = 0,\ \mu(\Omega) = 1$;
2. $A \subset B$ *implies* $\mu(A) \leq \mu(B)$.

From now on, the latter definition will be considered.

3 Relations among fuzzy measures and divergence measures

Proposition 2 *Let D be a divergence measure such that $D(\Omega, \emptyset) = 1$. Then the mapping $\mu : \widetilde{\mathcal{P}}(\Omega) \mapsto [0,1]$ defined by*

$$\mu(A) = D(A, \emptyset)$$

is a fuzzy measure.

Proof.

1. $\mu(\emptyset) = D(\emptyset, \emptyset) = 0$ and $\mu(\Omega) = D(\Omega, \emptyset) = 1$.
2. If $A \subset B$, then it follows with property d3 that

$$\mu(A) = D(A, \emptyset) = D(B \cap A, \emptyset \cap A) \leq D(B, \emptyset) = \mu(B). \quad \square$$

Proposition 3 *Let $\Omega = \{x_1, \ldots, x_n\}$. If D is a local divergence measure such that $D(\Omega, \emptyset) = 1$, then the mapping $\mu : \widetilde{\mathcal{P}}(\Omega) \mapsto [0,1]$ defined by*

$$\mu(A) = D(A, \emptyset)$$

is an additive fuzzy measure. Moreover, if $A \cap \Omega^i(x_i) = B \cap \Omega^j(x_j)$ then it holds that

$$\mu(A \cap \Omega^i) = \mu(B \cap \Omega^j).$$

Proof. Let us assume that $A \cap B = \emptyset$. Denote $\Omega_a = \{x \in \Omega \mid A(x) > 0\}$, $\Omega_b = \{x \in \Omega \mid B(x) > 0\}$ and $\Omega_c = \{x \in \Omega \mid A(x) = B(x) = 0\}$. Since D is local, there exists a mapping h such that

$$\begin{aligned}
\mu(A \cup B) = D(A \cup B, \emptyset) &= \sum_{i=1}^{n} h(\max\{A(x_i), B(x_i)\}, 0) \\
&= \sum_{x \in \Omega_a} h(A(x), 0) + \sum_{x \in \Omega_b} h(B(x), 0) + \sum_{x \in \Omega_c} h(0, 0) \\
&= D(A, \emptyset) + D(B, \emptyset) = \mu(A) + \mu(B)
\end{aligned}$$

Therefore, μ is additive. Furthermore, if $A(x_i) = B(x_j)$, then

$$\mu(A \cap \Omega^i) = D(A \cap \Omega^i, \emptyset) = h(A(x_i), 0) = h(B(x_j), 0) = D(B \cap \Omega^j, \emptyset) = \mu(B \cap \Omega^j). \quad \square$$

Proposition 4 *Let μ be an additive fuzzy measure, then the mapping $D : \widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \mapsto \mathbb{R}$ defined by*

$$D_\mu(A, B) = \mu(A \cup B) - \mu(A \cap B)$$

is a normal divergence measure.

Proof. D_μ must satisfy axioms d1–d5. Obviously, $D_\mu(A,B) = \mu(A \cup B) - \mu(A \cap B) = D_\mu(B,A)$ and $D_\mu(A,A) = \mu(A \cup A) - \mu(A \cap A) = \mu(A) - \mu(A) = 0$.

Next, let us prove that $D_\mu(A \cup C, B \cup C) \le D_\mu(A,B)$ (a similar proof holds for $D_\mu(A \cap C, B \cap C) \le D_\mu(A,B)$). Let us decompose Ω in three disjoint subsets:

$$\Omega_1 = \{x \in \Omega \mid \max\{A(x), B(x)\} \le C(x)\}$$
$$\Omega_2 = \{x \in \Omega \mid \min\{A(x), B(x)\} \le C(x) < \max\{A(x), B(x)\}\}$$
$$\Omega_3 = \{x \in \Omega \mid C(x) < \min\{A(x), B(x)\}\}.$$

We now introduce the notation $H_i = H \cap \Omega_i$. It then holds that

$$((A \cup C) \cup (B \cup C))_1 = C_1$$
$$((A \cup C) \cap (B \cup C))_1 = C_1$$
$$((A \cup C) \cup (B \cup C))_2 = (A \cup B)_2$$
$$((A \cup C) \cap (B \cup C))_2 = C_2$$
$$((A \cup C) \cup (B \cup C))_3 = (A \cup B)_3$$
$$((A \cup C) \cap (B \cup C))_3 = (A \cap B)_3.$$

The additivity of μ then implies that

$$\begin{aligned}
&D_\mu(A \cup C, B \cup C)\\
&= \mu((A \cup C) \cup (B \cup C)) - \mu((A \cup C) \cap (B \cup C))\\
&= \mu(((A \cup C) \cup (B \cup C))_1) + \mu(((A \cup C) \cup (B \cup C))_2) + \mu((A \cup C) \cup (B \cup C))_3)\\
&\quad -\mu(((A \cup C) \cap (B \cup C))_1) - \mu(((A \cup C) \cap (B \cup C))_2) - \mu(((A \cup C) \cap (B \cup C))_3)\\
&= \mu((A \cup B)_2) + \mu((A \cup B)_3) - \mu(C_2) - \mu((A \cap B)_3).
\end{aligned}$$

The monotonicity of μ implies that $\mu((A \cup B)_1) \ge \mu((A \cap B)_1)$ and $\mu(C_2) \ge \mu((A \cap B)_2)$. We then continue

$$\begin{aligned}
D_\mu(A \cup C, B \cup C) &\le \mu((A \cup B)_1) + \mu((A \cup B)_2) + \mu((A \cup B)_3)\\
&\qquad -\mu((A \cap B)_1) - \mu((A \cap B))_2 - \mu((A \cap B)_3)\\
&= \mu(A \cup B) - \mu(A \cap B) = D(A,B).
\end{aligned}$$

Therefore, D_μ is a divergence measure. Axioms d4 and d5 can be proven easily:

$$D_\mu(\Omega, \emptyset) = \mu(\Omega \cup \emptyset) - \mu(\Omega \cap \emptyset) = \mu(\Omega) - \mu(\emptyset) = 1$$

and

$$D_\mu(A \cup B, A \cap B) = \mu(A \cup B) - \mu(A \cap B) = D_\mu(A,B). \quad \square$$

Proposition 5 *Let $\Omega = \{x_1, x_2, \ldots, x_n\}$ and let μ be an additive fuzzy measure such that $A \cap \Omega^i(x_i) = B \cap \Omega^j(x_j)$ implies that*

$$\mu(A \cap \Omega^i) = \mu(B \cap \Omega^j).$$

Then D_μ is a local divergence measure.

Proof. Since μ is additive and $\Omega^i \cap \Omega^j = \emptyset$ for $i \neq j$, it holds that

$$\begin{aligned} D_\mu(A,B) - & D_\mu(A \cup \Omega^i, B \cup \Omega^i) \\ &= \mu(A \cup B) - \mu(A \cap B) - \mu((A \cup B) \cup \Omega^i) + \mu((A \cap B) \cup \Omega^i) \\ &= \sum_{j=1}^{n} \mu((A \cup B) \cap \Omega^j) - \sum_{j=1}^{n} \mu(A \cap B \cap \Omega^j) \\ &\quad - \sum_{j=1}^{n} \mu((A \cup B \cup \Omega^i) \cap \Omega^j) + \sum_{j=1}^{n} \mu(((A \cap B) \cup \Omega^i) \cap \Omega^j). \end{aligned}$$

If $i \neq j$, then $(A \cup B \cup \Omega^i) \cap \Omega^j = (A \cup B) \cap \Omega^j$ and $((A \cap B) \cup \Omega^i) \cap \Omega^j = A \cap B \cap \Omega^j$. If $i = j$, then $(A \cup B \cup \Omega^i) \cap \Omega^j = \Omega^i$ and $((A \cap B) \cup \Omega^i) \cap \Omega^j = \Omega^i$. Applying the above, it follows that

$$D_\mu(A,B) - D_\mu(A \cup \Omega^i, B \cup \Omega^i) = \mu((A \cup B) \cap \Omega^i)) - \mu(A \cap B \cap \Omega^i).$$

Let $g(A(x_i)) = \mu(A \cap \Omega^i)$. If $A(x_i) = B(x_j)$, then

$$g(A(x_i)) = \mu(A \cap \Omega^i) = \mu(B \cap \Omega^j) = g(B(x_j))$$

and g is well defined. Let h be defined by

$$h(A(x_i), B(x_i)) = g(\max\{A(x_i), B(x_i)\}) - g(\min\{A(x_i), B(x_i)\}).$$

Thus it holds that

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

and D_μ is local. $\square$

Theorem 1 *Let μ be a fuzzy measure. Then there exists a divergence measure D such that*

$$\mu(A) = D(A, \emptyset).$$

Proof. Let us define

$$N_A(x) = \begin{cases} 1 & \text{if } A(x) = 0 \\ 0 & \text{if } A(x) > 0 \end{cases}$$

$$D(A,B) = \begin{cases} 0 & \text{if } A \subseteq B \text{ and } B \subseteq A \\ \mu(B \cap N_A) & \text{if } A \subseteq B \text{ and } B \not\subseteq A \\ \mu(A \cap N_B) & \text{if } A \not\subseteq B \text{ and } B \subseteq A \\ 1 & \text{if } A \not\subseteq B \text{ and } B \not\subseteq A \end{cases}.$$

D must satisfy axioms d1–d3 to be a divergence measure. Obviously, $D(A,B) = D(B,A)$ and $D(A,A) = 0$. Next, let us prove that $D(A\cup C, B\cup C) \leq D(A,B)$ and $D(A\cap C, B\cap C) \leq D(A,B)$. By definition, we have that

$$D(A\cup C, B\cup C) = \begin{cases} 0 & \text{if } A\cup C \subseteq B\cup C \text{ and } B\cup C \subseteq A\cup C \\ \mu((B\cup C)\cap N_{A\cup C}) & \text{if } A\cup C \subseteq B\cup C \text{ and } B\cup C \not\subseteq A\cup C \\ \mu((A\cup C)\cap N_{B\cup C}) & \text{if } A\cup C \not\subseteq B\cup C \text{ and } B\cup C \subseteq A\cup C \\ 1 & \text{if } A\cup C \not\subseteq B\cup C \text{ and } B\cup C \not\subseteq A\cup C \end{cases}$$

and

$$D(A\cap C, B\cap C) = \begin{cases} 0 & \text{if } A\cap C \subseteq B\cap C \text{ and } B\cap C \subseteq A\cap C \\ \mu((B\cap C)\cap N_{A\cap C}) & \text{if } A\cap C \subseteq B\cap C \text{ and } B\cap C \not\subseteq A\cap C \\ \mu((A\cap C)\cap N_{B\cap C}) & \text{if } A\cap C \not\subseteq B\cap C \text{ and } B\cap C \subseteq A\cap C \\ 1 & \text{if } A\cap C \not\subseteq B\cap C \text{ and } B\cap C \not\subseteq A\cap C. \end{cases}$$

If $B\cup C \subseteq A\cup C$ and $A\cup C \subseteq B\cup C$, or, $B\cap C \subseteq A\cap C$ and $A\cap C \subseteq B\cap C$, then it is obvious that $D(A\cup C, B\cup C) = 0 \leq D(A,B)$ or $D(A\cap C, B\cap C) = 0 \leq D(A,B)$, respectively.

If $B\cup C \not\subseteq A\cup C$ and $A\cup C \not\subseteq B\cup C$, or, $B\cap C \not\subseteq A\cap C$ and $A\cap C \not\subseteq B\cap C$, then it is obvious that $A \not\subseteq B$ and $B \not\subseteq A$. Then $D(A\cup C, B\cup C) = 1 = D(A,B)$ or $D(A\cap C, B\cap C) = 1 = D(A,B)$, respectively.

If $B\cup C \not\subseteq A\cup C$ or $B\cap C \not\subseteq A\cap C$, then $B \not\subseteq A$. In this case, if $A \not\subseteq B$, then $D(A,B) = 1$. Therefore, $D(A\cup C, B\cup C) \leq 1 = D(A,B)$ or $D(A\cap C, B\cap C) \leq 1 = D(A,B)$, respectively. If $B \not\subseteq A$ and $A \subseteq B$, then $D(A,B) = \mu(B\cap N_A)$ and $D(A\cup C, B\cup C) = \mu((B\cup C)\cap N_{A\cup C})$ or $D(A\cap C, B\cap C) = \mu((B\cap C)\cap N_{A\cap C})$, respectively.

Since

$$\begin{aligned} (B\cup C)\cap N_{A\cup C}(x) &= \begin{cases} B(x) & \text{if } A(x) = 0 \text{ and } C(x) = 0 \\ 0 & \text{if } A(x) = 0 \text{ and } C(x) > 0 \\ 0 & \text{if } A(x) > 0 \end{cases} \\ (B\cap C)\cap N_{A\cap C}(x) &= \begin{cases} \min\{B(x), C(x)\} & \text{if } A(x) = 0 \\ 0 & \text{if } A(x) > 0 \end{cases} \\ B\cap N_A(x) &= \begin{cases} B(x) & \text{if } A(x) = 0 \\ 0 & \text{if } A(x) > 0 \end{cases} \end{aligned}$$

it follows that $(B\cup C)\cap N_{A\cup C} \subseteq B\cap N_A$ and $(B\cap C)\cap N_{A\cap C} \subseteq B\cap N_A$. Hence $\mu((B\cup C)\cap N_{A\cup C}) \leq \mu(B\cap N_A)$ and $\mu((B\cap C)\cap N_{A\cap C}) \leq \mu(B\cap N_A)$. Therefore, axiom d3 holds. □

4 Probabilistic divergence measure for the finite case

Now, let us turn to another problem: let D be a divergence measure and suppose that P is a probability distribution over $\Omega = \{x_1, \ldots, x_n\}$. We want to define D_P such that D_P remains a divergence measure, but it also uses the information given by P. Note that, intuitively, if P is the uniform distribution, it makes sense that $D_P = D$, i.e. the divergence measure remains the same when we have no information. Thus, we have added the following axiom:

d6. $D_P(A, B) = D(A, B)$ if P represents the uniform distribution.

In fact, the idea is to correct the initial divergence measure into the new one, in which we use the probability distribution and that is a better one in the sense that more information is used. Thus, we propose the following definition:

Definition 6 *Let D be a divergence measure and P be a probability distribution over Ω. We define the* **probabilistic divergence measure** *(or probabilistic divergence) associated to D and P as:*

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

where parentheses for probabilities mean a permutation such that $p_{(1)} \geq p_{(2)} \geq ... \geq p_{(n+1)} = 0$ and $A_{(i)}(x_j) = \min\{A(x_j), \Omega_i(x_j)\}$ with

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

(and similarly for $B_{(i)}$). $\mathcal{E}(P)$ is the information energy [11], which is defined as

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

Subset $A_{(i)}$ means that we are only interested in the coordinates that have at least a probability of $p_{(i)}$ and thus we try to take out the importance of the other coordinates. This idea is easier to see for local measures as we prove further on.

Finally, $\mathcal{E}(P)$ is necessary in order to verify axiom d6 as is proved below. We could consider other values: for example, for the finite case we could multiply by n or divide by $p_{(1)}$. However, the first option is no longer valid in the countable case and with the second one we obtain

$$D_P(A, B) \leq D(A, B),$$

causing the idea of correcting the values of $D(A, B)$ to be a little bit hidden.

Proposition 6 *D_P satisfies axioms* **d1**, **d2**, **d3** *and* **d6**.

Proof.

1. Axiom d1 follows easily:

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

2. Similarly for axiom d2:

$$D_P(A,A) = \frac{1}{\mathcal{E}(P)}\sum_{i=1}^{n}(p_{(i)} - p_{(i+1)})D(A_{(i)}, A_{(i)})$$
$$= \frac{1}{\mathcal{E}(P)}\sum_{i=1}^{n}(p_{(i)} - p_{(i+1)})0 = 0.$$

3. For axiom d3, we prove for instance $D(A \cup C, B \cup C) \leq D(A,B)$:

$$D(A \cup C, B \cup C) = \frac{1}{\mathcal{E}(P)}\sum_{i=1}^{n}(p_{(i)} - p_{(i+1)})D((A \cup C)_{(i)}, (B \cup C)_{(i)})$$
$$\leq \frac{1}{\mathcal{E}(P)}\sum_{i=1}^{n}(p_{(i)} - p_{(i+1)})D(A_{(i)}, B_{(i)}) = D(A,B).$$

4. Finally, we show that axiom d6 holds. If $p(x_i) = \frac{1}{n}$, then all summands vanish except the last one. Hence,

$$D_P(A,B) = \frac{1}{\mathcal{E}(P)}\frac{1}{n}D(A,B) = D(A,B)$$

since $\frac{1}{\mathcal{E}(P)} = n$ in the uniform case. □

Let us now see some properties of this measure:

Proposition 7 *Let D be a local divergence measure. Then D_P can be written as follows:*

$$D_P(A,B) = \frac{1}{\mathcal{E}(P)}\sum_{i=1}^{n} p(x_i)h(A(x_i), B(x_i)).$$

Proof. For a local divergence measure it holds that

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

Let us consider x_i and suppose that $p(x_i) = p_{(j)}$. Then, if we apply this expression in the definition of D_P we obtain that $h(A(x_i), B(x_i))$ is multiplied by:

$$\frac{1}{\mathcal{E}(P)} \sum_{i=j}^{n} (p_{(i)} - p_{(i+1)}) = \frac{1}{\mathcal{E}(P)} (p_{(j)} - p_{(n+1)}) = \frac{1}{\mathcal{E}(P)} p_{(j)}$$

which concludes the proof. □

Proposition 8 *In general, for a local measure D, D_P is not a local measure, except for the uniform case.*

Proof. Suppose D_P is a local measure. Then there exists a mapping g such that

$$D_P(A, B) = \sum_{i=1}^{n} g(A(x_i), B(x_i)).$$

Using the previous proposition we obtain that

$$g(A(x_i), B(x_i)) = \frac{1}{\mathcal{E}(P)} p(x_i) h(A(x_i), B(x_i)).$$

Now, we can find a probability distribution P and two fuzzy subsets A and B such that $A(x_i) = A(x_j)$, $B(x_i) = B(x_j)$ and $p(x_i) \neq p(x_j)$. Then, $g(A(x_i), B(x_i)) \neq g(A(x_j), B(x_j))$ contradicting our hypothesis. □

Remark 1 *In fact, the only case where this holds, is the uniform case. This is true because the function h for D_P is different for each coordinate. However, each coordinate is independent of the others and the only difference is that the coordinates don't have the same weight. Note that this was expected as divergence measures are special cases of probabilistic divergence measures when all coordinates are equally important, i.e., when we have the uniform distribution.*

Remark 2 *For the local case, the values that the coordinates considered take are not important. It is only necessary that both subsets take the same value. Just take the expression for local measures and remark that $h(x, x) = 0, \forall x \in [0, 1]$. This result is not true for general divergence measures.*

Example 1 *Let us consider the local divergence measure D defined by*

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

Considering a probability distribution P, we obtain with Proposition 7 that

$$D_P(A, B) = \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{n} p(x_i) |A(x_i) - B(x_i)|,$$

which seems an intuitive way to extend D when we have a probability distribution.

Note that D_P can be greater than D, that it can be smaller than D or even they can take the same value. This is due to the fact that the probability measure is used to correct the value of D with the new information.

Remark 3 *Note that we can change $p(x_i)$ by a general weight $u(x_i)$. Then, all the results remain true and the probabilistic case is just a special case.*

5 Probabilistic divergence measure for a discrete distribution

In the countable case, we can always assure that the maximum of the probabilities exists and we can order the probabilities decreasingly. This leads us to propose the following definition:

Definition 7 *Let D be a divergence measure and P be a discrete probability distribution over Ω. We define the* **probabilistic divergence measure** *associated to D and P as:*

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

where parentheses for probabilities mean a permutation such that $p_{(1)} \geq p_{(2)} \geq \ldots$ and $A_{(i)}(x_j) = \min\{A(x_j), \Omega_i(x_j)\}$ with

$$\Omega_i(x_j) = \begin{cases} 1 & \text{if } \; p(x_j) \geq p_{(i)} \\ 0 & \text{if } \; p(x_j) < p_{(i)} \end{cases} \tag{2}$$

(and similarly for $B_{(i)}$).

This definition is very similar to the one for the finite case. Note also that the information energy is always a finite value.

Proposition 9 $D_P(A,B) \leq \frac{1}{\mathcal{E}(P)} D(A,B)$.

Proof. As $A_{(i)} = A \cap \Omega_i$ and $B_i = B \cap \Omega_i$, it follows with axiom **d3** that $D(A_{(i)}, B_{(i)}) \leq D(A,B)$. Hence,

$$\begin{aligned} D_P(A,B) &= \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{\infty} (p_{(i)} - p_{(i+1)}) D(A_{(i)}, B_{(i)}) \\ &\leq \frac{1}{\mathcal{E}(P)} D(A,B) \sum_{i=1}^{\infty} (p_{(i)} - p_{(i+1)}) \leq D(A,B). \quad \square \end{aligned}$$

Corollary 1 *If $D(A,B) < \infty$, then $D_P(A,B) < \infty$.*

Example 2 *Consider again the local divergence measure D defined by*

$$D(A,B) = \sum_{i=1}^{\infty} |A(x_i) - B(x_i)|.$$

Then, considering a probability distribution P an easy calculation leads to the following expression:

$$D_P(A,B) = \frac{1}{\mathcal{E}(P)} \sum_{i=1}^{\infty} p(x_i) |A(x_i) - B(x_i)|.$$

	Gold	Bond	Stock	C-D	∅
Gold	0				2,62
Bond	0,91	0			2,79
Stock	1,39	0,47	0		2,81
C-D	0,62	1,10	0,75	0	2,08

Table 3: Probabilistic Divergence I

6 Application

Let us suppose, in order to establish a relation with the utility approach and the expected utility criterion, that we choose the divergence measure given by

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

Thus,

$$D_P(A,B) = \frac{1}{\sum_{i=1}^{n} p(x_i)^2} \sum_{i=1}^{n} p(x_i)|A(x_i) - B(x_i)|.$$

Of course, we could have considered another divergence measure. We consider this one because we obtain the expected utility divided by the information energy.

We consider again the example from the introduction. Note that the worst action is the one with utility values 0 for all possible states of the world (the empty set). Thus, the best one will be the one with the highest divergence value with the empty set. The values obtained are 2,62 2,79 2,81 and 2,08 respectively. Note also that we would have obtained the same result if we would have taken the universe as the best option and thus the best action would be the one with the smallest divergence value with the universe. We can see that bond and stock take almost the same value (2,79 and 2,81).

Now, let us compute the divergence table between pairs of actions using the probability values given in the introduction. The results are given in Table 3. From this table, we can see that the divergence measure between bond and stock is very small. That means that if we change slightly the probability values or the utility values (usually these values are approximations based on experimental results or even subjective values), the divergence value between bond and stock would remain very small, and thus there is not a big difference between these two possibilities in order to take a decision.

However, let us suppose that we had another probability distribution. Let us suppose that we have replaced Table 1 by Table 4.

In this case, gold and bond have very similar expected utility values. However, if we compute the divergence table (Table 5), we see that the divergence between gold and bond is rather big. That means that a small change in the probability values can lead to a big difference between the expected values. In fact, note that

	States of Nature				
	Large Rise	Small Rise	No Change	Small Fall	Large Fall
Probability	0.30	0.05	0.05	0.10	0.50

Table 4: Probability II

	Gold	Bond	Stock	C-D	$\emptyset$
Gold	0				0,515
Bond	0,3125	0			0,5125
Stock	0,525	0,2125	0		0,4
C-D	0,145	0,2225	0,43	0	0,5

Table 5: Probabilistic Divergence II

the utility values of these two actions are rather different. The difference with the other case is that there is not a really important change by choosing bond instead of gold as they are very similar; however, in this new case, if our values are not right we can fail in our decision. Thus, it could be useful to make another study of our utility values and the probability distribution in order to assure that our decision is really the best.

7 Conclusions

Divergence measures are able to give a degree of evidence or belief of an action A being (not) similar to B. The expected utility criterion we have interpreted is a projection of a divergence measure. Our interpretation says that for the expected utility criterion, the degree of evidence or belief for each possible payoff, is the divergence measure between A and $\emptyset$, the least preferred outcome.

As we have exposed another viewpoint on fuzzy measures based on divergence measures, we should develop a theory of divergence measures on the basis of this generalization. We have given some properties of divergence measures establishing a relationship between divergence measures and fuzzy measures, in particular when the divergence measures are local and the fuzzy measures are additive.

We have also introduced the concept of a probabilistic divergence. It allows us to correct the original divergence measures when probabilistic information is available. In the example, we have seen that the probability measure is rather important in order to choose the best action and also to study the differences between a pair of actions. This study can be interesting in order to decide if the best actions are rather similar and hence it does not care which one must be chosen or, on the other hand, if they are very different and thus a more deep study is needed in order to fix the values.

An interesting problem is the extension of this concept to the continuous case. A possible definition could be

$$D_f(A,B) = \frac{1}{\mathcal{E}(f)} \int_0^{\infty} D(A_{(\alpha)}, B_{(\alpha)}) d\alpha,$$

where

$$A_{(\alpha)}(x) = \begin{cases} 1 & \text{if} \quad f(x) \geq \alpha \\ 0 & \text{if} \quad f(x) < \alpha \end{cases} .$$

However, in this case we can not assure that $\mathcal{E}(f)$ is a finite value or that the integral value is finite. In order to study the properties and advantages and disadvantages of this definition much work remains undone.

Acknowledgement

The authors want to thank Dr. S. Montes for her help.

References

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

[2] B. Burbea and C. Rao, *Entropy differential metric, distance and divergence measures in probability spaces: a unified approach*, J. Multivariate Anal. **12** (1982), 575–596.

[3] B. De Baets and R. Mesiar, *T-partitions*, Fuzzy Sets and Systems **97** (1998), 213–223.

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

[5] M. Grabisch, *The application of fuzzy integrals in multicriteria decision making*, European J. Oper. Res. **89** (1996), 145–456.

[6] J. Kapur, *Measures of informations and their applications*, John Wiley, 1994.

[7] J. Keller and R. Krishnapuram, *Fuzzy decision Models in Computer Vision*, in: Fuzzy sets, neuronal networks and soft computing (R. Yager and L. Zadeh, eds.), Van Nostrand Reinhold, 1994.

[8] J. Lawrence and B. Pasternack, *Applied Management Science*, John Wiley, 1998.

[9] S. Montes, *Particiones y medidas de divergencia en modelos difusos*, Ph.D. Thesis, University of Oviedo, 1998.

[10] S. Montes, P. Gil and C. Bertoluzza, *Divergence between fuzzy sets and fuzziness*, Proceedings IPMU '98, Paris, 1998, pp. 1030–1035.

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

[12] L. Pardo, *Teoría de la Información Estadística*, Ed. Hespérides, 1997.

[13] M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. thesis, Tokyo Institute of Technology, 1974.