INFORMATION AND CONTROL **24**, 55–73 (1974)

Entropy of L-Fuzzy Sets*

ALDO DE LUCA AND SETTIMO TERMINI

Laboratorio di Cibernetica del Consiglio Nazionale delle Ricerche, Arco Felice, Napoli, Italy

The notion of "entropy" of a fuzzy set, introduced in a previous paper in the case of generalized characteristic functions whose range is the interval [0, 1] of the real line, is extended to the case of maps whose range is a poset L (or, in particular, a lattice).

Some of the reasons giving rise to the non-comparability of the truth values and then the necessity of considering poset structures as range of the maps are discussed.

The interpretative problems of the given mathematical definitions regarding the connections with decision theory are briefly analyzed.

1. INTRODUCTION

In this work the notion of entropy of a fuzzy set (Zadeh, 1965) introduced in De Luca and Termini (1972) will be extended to the case of L-fuzzy sets (Goguen, 1967), i.e., maps from a given set I to a partly ordered set (poset) L. The difference between the case in which L coincides with the interval [0, 1] of the real line (fuzzy sets) and that considered here, in which L is a general poset, is similar to the one existing between a *multivalued logic* (see, for instance, Łukasiewicz and Tarski, 1930) and a logic in which the truth-values of the propositions are not always comparable (see, for instance, Koopman, 1940).

In the next section we shall collect some mathematical preliminaries that will be used in the following.

Before introducing the formal definition of entropy of an L-fuzzy set, we shall, then, briefly discuss how and why the necessity arises of considering poset structures as range of the generalized characteristic functions. The mathematical definition of the entropy of an L-fuzzy set will then be given in the general case. This entropy is a matrix quantity so that the L-fuzzy sets cannot always be compared by means of their entropies. A more detailed

* An abstract of this paper was published in *Notices A.M.S.* **19**, A710 (1972).

discussion on the appearance of the non-comparability of the truth-values for systems described by means of "independent" and "measurable" properties is made. The interpretation of the new mathematical concepts introduced will then be presented, taking into particular account its connections with decision theory. In this last case a (natural) direct interpretation of the entropy as a measure of the uncertainty in decision taking is given.

We shall finally indicate what are, in our opinion, some possible implications of the previous theory, stressing at the same time its main features.

2. Mathematical Preliminaries

In this section we list some definitions and mathematical results that will be of use.

Let I be a set and L a poset.

Definition 2.1. *An L-fuzzy set defined on I is any map from I to L* (Goguen, 1967).

If the poset L coincides with the interval [0, 1] of the real line we obtain the definition of fuzzy set as introduced by Zadeh (1965).

We denote by $\mathscr{L}(I, L)$ the class of all the L-fuzzy sets defined on I; the class of all the fuzzy sets defined on I will be simply defined by $\mathscr{L}(I)$.

We remember that any closed operation defined on L can be induced point-by-point on $\mathscr{L}(I, L)$. In particular, if L is a lattice then one can give to $\mathscr{L}(I, L)$ a lattice structure by means of the binary operations $\vee$ and $\wedge$ defined, for any pair f and g of elements of $\mathscr{L}(I, L)$, as:

$$\begin{aligned} (f \vee g)(x) &\equiv \text{l.u.b.}\{f(x), g(x)\}, \\ (f \wedge g)(x) &\equiv \text{g.l.b.}\{f(x), g(x)\}, \end{aligned} \tag{2.1}$$

for all $x \in I$, where l.u.b. and g.l.b. denote the *least upper bound* and the *greatest lower bound*, respectively.

Some remarks on the algebraic structure of $\mathscr{L}(I)$ can be found in De Luca and Termini (1970). We remember, in particular, that $\mathscr{L}(I)$ is a non-complemented lattice with respect to the operations (2.1).

Definition 2.2. *Let I be a finite set. An "entropy" measure on $\mathscr{L}(I)$ is a non-negative functional*

$$d\colon \mathscr{L}(I) \to \mathbf{R}, \tag{2.2}$$

R *denoting the set of non-negative real numbers, such that the following properties are satisfied:*

P_1. $d(f) = 0$ *if and only if f is a classical characteristic function.*

P_2. $d(f)$ *is maximum if and only if* $f(x) = 1/2$ *for all* $x \in I$.

P_3. $d(f^*) \leqslant d(f)$, *where f^* is any "sharpened" version of f, that is,* $f^*(x) \leqslant f(x)$ *for* $f(x) \leqslant 1/2$ *and* $f^*(x) \geqslant f(x)$ *for* $f(x) \geqslant 1/2$.

The previous properties, as discussed in De Luca and Termini (1972), are the weakest requirements that, intuitively, a measure of the "degree of fuzziness" of f must satisfy. In De Luca and Termini (1972), in analogy with Shannon's entropy of information theory, we fixed a particular measure of entropy given by the following functional

$$k \sum_{i=1}^{N} S(f_i), \tag{2.3}$$

where $f_i = f(x_i)$, $N = \#I$, $S(x)$ is the *Shannon function*

$$S(x) = -x \ln x - (1 - x) \ln(1 - x),$$

and k a positive constant. The functional (2.3) besides the properties P_1, P_2, P_3 satisfies many other properties.

In Capocelli and De Luca (1972) a large class of functionals satisfying P_1, P_2, P_3, to which (2.3) belongs, has been taken into account. A functional of the class is given by

$$d(f) \equiv \sum_{i=1}^{N} T(f_i), \tag{2.4}$$

where $T(x) \equiv \mu(x) + \mu(1 - x)$, μ being any continuous and strictly concave function in the interval (0, 1) such that

$$\lim_{x \to 0} \mu(x) = \lim_{x \to 1} \mu(x) = 0.$$

Entropy measures as (2.4) have been considered by Vajda (1969) in problems of statistical pattern recognition.

In the following we shall call (2.3) "logarithmic entropy" of a fuzzy set using the word "entropy" for any arbitrary functional (2.4) with a T not specified.

From the fact that μ is a concave function it follows that

$$d(f) \leqslant NT(P(f)/N), \tag{2.5}$$

where $P(f)$ is the *power* of f, i.e., $P(f) \equiv \sum_{i=1}^{N} f_i$. A further property of the entropy d that we shall use in the following is that *d is a non-negative valuation on the lattice $\mathscr{L}(I)$*, that is,

$$d(f) + d(g) = d(f \vee g) + d(f \wedge g), \qquad \text{for all} \quad f, g \in \mathscr{L}(I). \tag{2.6}$$

3. Structure of the Range of the Generalized Characteristic Functions

In this section we briefly stress how the use of a poset L (in particular a lattice) as the range of generalized characteristic functions arises in a natural way when dealing with the description of sets of objects by means of more than a single property.

We shall first see that it is possible to attain the notion of L-fuzzy set in two ways; a unique formal scheme, including them as particular cases, will then be presented.

The interpretation of the formalism will be given in the following sections.

Let us consider a set I of objects and M properties $P^j (j = 1,..., M)$; suppose that M fuzzy sets

$$f^j: I \to [0, 1] \qquad (j = 1,..., M) \tag{3.1}$$

are given, $f^j(x)$ being interpreted as a numerical evaluation of the "degree" to which x enjoys the property P^j.

It is then possible to associate a matrix $M \times 1$, whose elements are the degrees with which x enjoys the M properties, to any given element $x \in I$:

$$x \to \begin{bmatrix} f^1(x) \\ \vdots \\ f^M(x) \end{bmatrix}. \tag{3.2}$$

A map f from I to the set L of all $M \times 1$ matrices whose elements take values in [0, 1] is then, in such a way, constructed.

L is a lattice with respect to the partial order relation defined as

$$z \leqslant y \Leftrightarrow z_j \leqslant y_j, \qquad \text{for all} \quad j \in J, \tag{3.3}$$

where $J \equiv \{1,..., M\}$ and z and y are two elements of L of components z_j and y_j, $j \in J$. L has a greatest and a least element given by the $M \times 1$ matrices whose components are all ones and all zeros, respectively.

In such a way a set of M maps (3.1) naturally determines an L-fuzzy set.

Let us now consider the limiting case in which the objects of I are described by means of only one property of a given ensemble. This situation occurs, for instance, if our set I is such that for each object one property is more meaningful than the others, or when only one property for each object is available.

We obtain, in this way, a partition of I in some classes A_j, equal in number to the properties taken into consideration.

It is, obviously, possible to compare the "degrees" $f^j(x)$ with which the various elements of any A_j enjoy the corresponding property P^j; these "degrees" may be, then, ordered in a chain. It is meaningless, however, to compare "degrees" referring, for instance, to different "independent" properties (see Section 5); in this case a collection of disjoint chains, one for each property, is obtained.

If $f^j(x)$ denotes for each property P^j, $j \in J$, the "degree" to which the element $x \in A_j$ enjoys P^j, one may equivalently give a *map* f from I to a suitable set L and a map

$$\alpha: L \to [0, 1] \tag{3.4}$$

such that $\alpha(f(x)) = f^j(x)$.

L has to be constructed taking into account all the allowed partitions of I and all the possible degrees $f^j(x)$, $j \in J$.

Moreover, one imposes that if y and z are two elements of L such that $\alpha(y) = \alpha(z)$ then $y = z$.

L can be partly ordered with respect to the partial order relation

$$f(x) \leqslant g(y) \Leftrightarrow \{\alpha(f(x)) \leqslant \alpha(g(y)) \text{ and } xCy\}, \tag{3.5}$$

where xCy means that x and y belong to the same element of a partition of I. The previous examples show that the use of a poset as the range of generalized characteristic functions is not a purely mathematical extension but naturally comes out when one describes a certain class of objects by means of more than one property.

We now give a general mathematical scheme that includes the previous cases and a more general one in which, for each object x of I, a variable number of properties chosen among M fixed ones is considered. The main difference with the cases considered above is that we will start from a purely formal scheme, allowing various interpretations. It will be possible also to introduce in a more direct way the formal notion of entropy. Detailed interpretative problems will be postponed to the following sections.

Let f be a map from I to a set L and

$$\alpha_j : D_j \to [0, 1], \qquad j \in J, \tag{3.6}$$

M maps with D_j subsets of L. In such a way, for any given f, one obtains a set of M maps:

$$\alpha_j \text{ of}: A_j(f) \to [0, 1], \qquad j \in J, \tag{3.7}$$

where $\alpha_j \circ f(x) \equiv \alpha_j(f(x))$ for $x \in A_j(f)$, being $A_j(f) \equiv f^{-1}(D_j)$ that is the *inverse image* of D_j with respect to f.

We denote by $R(y)$, for any $y \in L$, the subset of J defined as

$$R(y) \equiv \{j \mid j \in J \text{ and } y \in D_j\}. \tag{3.8}$$

We suppose the maps α_j to be such that $\alpha_j \circ f$, $j \in J$, uniquely determine f, that is, if $z, w \in L$ are such that $R(z) = R(w)$ and $\alpha_j(z) = \alpha_j(w)$, $j \in R(z)$, then $z = w$. Because of this condition we assume that, for all x and f, $R(f(x))$ is *not empty* so that the $A_j(f)$, $j \in J$, for any f, determine a *cover* of I.

Given these assumptions L can be partly ordered as

$$z \leqslant w \Leftrightarrow R(z) = R(w) \quad \text{and} \quad \alpha_j(z) \leqslant \alpha_j(w), \quad \text{for all} \quad j \in R(z). \tag{3.9}$$

If α_j of (x) coincides, for any $j \in J$, with the degree to which x enjoys a certain property P^j, then $A_j(f)$ is the subset of I for which the property P^j is taken into account. In the first example $A_j(f) \equiv I$, $j \in J$, and L is the lattice of all $M \times 1$ matrices with components in [0, 1]; in the second one the $A_j(f)$, for any f, form a *partition* of I: for any x only one of the maps α_j is defined, so that the set of maps α, , $j \in J$, acts as a "projection operator" on the single properties, one for any element. In this case L is a poset formed by M chains.

Finally, we note that in the ordinary definition of a fuzzy set the map that gives a numerical evaluation of the degree with which a property is enjoyed is implicitly given and assumed equal to the identity function.

4. Entropy of L-Fuzzy Sets

The formal notion of the entropy $d(f)$ of an L-fuzzy set f is now introduced starting from the general definition 2.1.

Let us consider L-fuzzy sets defined on a finite set I taking values in a general poset L. Let α_j, $j \in J$, be M maps

$$\alpha_j : D_j \to [0, 1], \qquad j \in J, \tag{4.1}$$

D_j being subsets of L. The maps α_j determine, for any fixed f, a set of M maps

$$\alpha_j \circ f\colon A_j(f) \to [0, 1], \qquad j \in J.$$

Let

$$\mathscr{L} \equiv \{\mathscr{L}(D) \mid D \subseteq L\},$$

where $\mathscr{L}(D)$ is the class of all the fuzzy sets defined on D. The set (4.1) of M maps α_j is an element of the cartesian product $\mathscr{L}^{(M)}$.

Let $\mathbf{R}$ the set of non-negative real numbers and d the entropy of an ordinary fuzzy set (see Definition 2.2). We give the following

DEFINITION 4.1. *The entropy of an L-fuzzy set f with respect to a set of maps α_j, $j \in J$, is the quantity $\mathscr{d}(f, \alpha)$ where $\mathscr{d}$ is the map*

$$\mathscr{d}\colon \mathscr{L}(I, L) \times \mathscr{L}^{(M)} \to \mathbb{R}^{(M)}, \tag{4.2}$$

satisfying the condition

$$d_j(f, \alpha) = d(\alpha_j \circ f), \qquad j \in J, \tag{4.3}$$

for all $f \in \mathscr{L}(I, L)$ and $\alpha \in \mathscr{L}^{(M)}$.

$d_j(f, \alpha)$ is, then, the entropy (see Definition 2.4) of the fuzzy set $\alpha_j \circ f$ defined on $A_j(f)$. $\mathscr{d}(f, \alpha)$ can be represented by a $M \times 1$ matrix

$$\mathscr{d}(f, \alpha) = \begin{bmatrix} d_1(f, \alpha) \\ \vdots \\ d_M(f, \alpha) \end{bmatrix}. \tag{4.4}$$

In the following, when no misunderstanding can arise, we shall drop the index α.

The set $\Delta \equiv \{\mathscr{d}(f, \alpha) \mid \alpha \in \mathscr{L}^{(M)}, f \in \mathscr{L}(I, L)\}$ may be partly ordered by means of the relation $\leqslant$ defined as

$$\mathscr{d}^a \leqslant \mathscr{d}^b \Leftrightarrow d_j^a \leqslant d_j^b, \qquad \text{for all} \quad j \in J, \tag{4.5}$$

where $\mathscr{d}^a$ and $\mathscr{d}^b$ are two elements of Δ.

It is easy to prove that Δ is a *lattice* with respect to the operations $\cup$ and $\cap$ induced by the order relation (4.5), consisting, for any pair of entropy matrices, in making the maximum and minimum of all the components. This lattice has a minimum element given by the matrix $\mathscr{d}_{\min} = 0$ and a maximum $\mathscr{d}_{\max}$ which is reached when $(\alpha_j \circ f)(x_i) = 1/2$ for all $x \in I$ and for all $j \in J$. We observe that Δ is closed under application of stochastic matrices to its elements.

From any entropy $\mathscr{d}(f)$ one can extract the scalar quantity $\delta(f)$ defined as

$$\delta(f) \equiv \sum_{j=1}^{M} d_j(f); \tag{4.6}$$

when $D_j = L$, $j \in J$, as in the first example of the previous section, (4.5) becomes

$$\delta(f) = \sum_{j=1}^{M} \sum_{i=1}^{N} T[(\alpha_j \circ f)(x_i)]. \tag{4.7}$$

If the D_j are disjoint, as in the second example, $\delta(f)$ reduces to

$$\delta(f) = \sum_{k=1}^{N} T[(\alpha_{j(k)} \circ f)(x_k)], \tag{4.8}$$

where $j(k)$ gives for any k the index of the unique map $\alpha_{j(k)}$ which is considered.

Let us now make, in view of the interpretation that will be given to the entropy $\mathscr{d}$, the following assumptions on the maps α_j :

A1. *The $A_j(f)$, $j \in J$, for any f, are a cover of I.*

A2. *For any pair (y, z) of elements of L one has*

$$y \leqslant z \Leftrightarrow R(y) \equiv R(z) \quad \text{and} \quad \alpha_j(y) \leqslant \alpha_j(z) \quad \text{with} \quad j \in R(y).$$

A3. *For any pair y and z of elements of L one has*

$$\alpha_j \circ (y \vee z) = (\alpha_j \circ y) \cup (\alpha_j \circ z),$$
$$\alpha_j \circ (y \wedge z) = (\alpha_j \circ y) \cap (\alpha_j \circ z), \qquad j \in J,$$

where $\cup$ and $\cap$ denote the lattice operations in $\mathscr{L}(I)$.

The examples considered in the previous section satisfy the assumptions A1 and A2; A3 is satisfied when $D_j = L$, $j \in J$.

If we suppose L to be a lattice and $D_j = L$, $j \in J$, one can prove under the hypothesis A2, A3 (A1 in this case is automatically satisfied) the following:

PROPOSITION 4.1. *The entropy $\mathscr{d}$ satisfies for any fixed $\alpha \in \mathscr{L}^M$ the property*

$$\mathscr{d}(f \vee g) + \mathscr{d}(f \wedge g) = \mathscr{d}(f) + \mathscr{d}(g) \tag{4.9}$$

that is, $\mathscr{d}$ is a (matrix) valuation on the lattice $\mathscr{L}(I, L)$.

To prove (4.8) one has to show that for all $j \in J$ and f and $g \in \mathscr{L}(I, L)$

$$d_j(f \vee g) + d_j(f \wedge g) = d_j(f) + d_j(g),$$

that is

$$d(\alpha_j \circ (f \vee g)) + d(\alpha_j \circ (f \wedge g)) = d(\alpha_j \circ f) + d(\alpha_j \circ g).$$

From the property (2.6) of valuation of d on the lattice $\mathscr{L}(I)$ one has

$$d(\alpha_j \circ f) + d(\alpha_j \circ g) = d((\alpha_j \circ f) \cup (\alpha_j \circ g)) + d((\alpha_j \circ f) \cap (\alpha_j \circ g)),$$

so that, under the assumption A3, (4.9) follows.

By means of procedure analogous to the one used in De Luca and Termini (1972) it is possible to introduce in the class $\mathscr{L}(I, L)$ of all the L-fuzzy sets defined on I the quotient set $\mathscr{L}(I, L)/\sim$ with respect to the equivalence relation

$$f \sim g \Leftrightarrow \mathscr{d}(f) = \mathscr{d}(g) \tag{4.10}$$

for a fixed $\alpha \in \mathscr{L}^{(M)}$.

To any element $K \in \mathscr{L}(I, L)/\sim$ one may associate the quantity $\mathscr{d}(K)$ defined as

$$\mathscr{d}(K) \equiv \mathscr{d}(f), \qquad \text{with} \qquad f \in K,$$

and partly order the equivalence classes by the relation

$$K_{\mathscr{d}_1} \leqslant K_{\mathscr{d}_2} \Leftrightarrow \mathscr{d}_1 \leqslant \mathscr{d}_2 .$$

We observe that, differently from what occurs in the case of fuzzy sets, the quotient set, in general, is not a chain.

DEFINITION 4.2. *The normalized entropy of an L-fuzzy set is the matrix*

$$\nu(f) \equiv \begin{bmatrix} \nu_1(f) \\ \vdots \\ \nu_M(f) \end{bmatrix}, \tag{4.11}$$

where

$$\nu_j(f) = \frac{1}{\# A_j(f)} d_j(f)$$

if $A_j(f) \neq \phi$ *and*

$$\nu_j(f) = \infty$$

if $A_j(f) = \phi$.

$\nu(f)$ is obviously a generalization of the normalized entropy of a fuzzy set (De Luca and Termini, 1972) and the component $\nu_j(f)$, $j \in J$, represents the average amount of entropy, connected with the property P^j, with regard to those elements of I for which P^j has been considered.

The normalized entropies ν can be partly ordered in a similar way to the entropies $\mathscr{d}$ (see Eq. 4.2); however, these order relations are not, in general, comparable.

5. Measurable and Independent Properties

Before dealing with the interpretation of the formal notion of entropy of an L-fuzzy set we take up again the interpretative problems related with the poset structure of the range of L-fuzzy sets. To this end the notions of measurable properties and independence of properties will be introduced more formally than was done in the previous sections.

Let a class U of objects be given and a property P defined on it. We suppose that the objects of U can enjoy to a different degree the property P. Mathematically a map

$$\psi_P : U \to [0, 1] \tag{5.1}$$

is given such that for all $x \in U$, $P(x) \leftrightarrow (\psi_P(x) = 1)$ and $\psi_P(x)$ is interpreted as *the degree to which x enjoys P* or, equivalently, *the degree of membership of x to the set*

$$P_1(U) \equiv \{x \in U \mid P(x)\}.$$

If $\psi_P(x) = 1$ we say that the property P is "present" in x; if $\psi_P(x) = 0$ we say that P is "absent" in x. A property P together with a map ψ_P given by (5.1) has been called *measurable* (Capocelli and De Luca, 1972) and, if $P_1(U) \not\equiv \phi$, *completely measurable.* ψ_P is called *measure* of P or *membership function.* The set $P_0(U) \equiv \{x \in U \mid \psi_P(x) = 0\}$ determines a further property $\sim P$ such that

$$(\sim P)(x) \leftrightarrow (\psi_P(x) = 0).$$

A noteworthy case is when, for all the objects of U,

$$\psi_P(x) + \psi_{\sim P}(x) = 1. \tag{5.2}$$

Two properties $P_1 \sim P$ completely measurable in U satisfying (5.2) are called *orthogonal.* Intuitively this is the case when the objects of U have "features intermediate" between those of $P_1(U)$ and $P_0(U)$.

Let us now consider M properties $P^1,\ldots, P^M$ defined in U of measures $\psi^1,\ldots, \psi^M$. Intuitively, the properties $P^j(j = 1 \cdots M)$ are *independent* if U is such that whatever object of U is considered the values of a certain number of measures do not determine the values of the others. To better clarify this concept let us denote by R_h the *range* of ψ^h:

$$R_h = \text{range}\, \psi^h \qquad (h = 1,\ldots, M), \tag{5.3}$$

and by $\psi_{i_1\cdots i_s}(x)$ the s-tuple $[\psi^{i_1}(x),\ldots, \psi^{i_s}(x)]$ with $i_1 < i_2 < \cdots < i_s$. We denote by $R_{i_1\cdots i_s}$ the range of $\psi_{i_1\cdots i_s}$.

Some definitions of "independence" of properties are now given, the first (Definition 5.1) is a formal definition of independence of two sets of properties; the second (Definition 5.2) introduces the notion of a set of independent properties.

DEFINITION 5.1. *Let $\mathscr{P} \equiv \{P^1,\ldots, P^M\}$ be a set of M measurable properties defined in U and two subsets $\mathscr{Q} \equiv \{P^{h_1},\ldots, P^{h_s}\}$ and $\mathscr{R} \equiv \{P^{k_1},\ldots, P^{k_r}\}$ of $\mathscr{P}$. $\mathscr{Q}$ and $\mathscr{R}$ are independent in U if and only if*

$$R_{h_1\cdots h_s k_1\cdots k_r} = R_{h_1\cdots h_s} \times R_{k_1\cdots k_r}, \tag{5.4}$$

where $\times$ denotes the Cartesian product of sets.

DEFINITION 5.2. *The measurable properties $P^1,\ldots, P^M$ are independent in U if and only if any $\{P^j\}(j = 1,\ldots, M)$ is independent of any set of other properties.*

By this definition and from Definition 5.1 a necessary and sufficient condition for a set of properties to be independent is obtained.

PROPOSITION 5.1. *M measurable properties $P^1,\ldots, P^M$ are independent if and only if*

$$R_{j_1\cdots j_s} = R_{j_1} \times \cdots \times R_{j_s} \tag{5.5}$$

for any s by s (simple) combination $j_1,\ldots, j_s$ of the indices $1,\ldots, M$.[1]

[1] *We note that even if Eq. (5.5) is not related to statistical considerations it looks similar to the conditions of "statistical independence" of a set of M events $E_1,\ldots, E_M$ if $R_{j_1\cdots j_s}$ is formally replaced by the "probability" of the event $E_{j_1} \cap \cdots \cap E_{j_s}$, R_{j_k} by the probability of E_{j_k}, and the Cartesian product by the ordinary product.*

If condition (5.5) is not satisfied, the properties $P^j(j = 1 \cdots M)$ will be called *dependent*. A pair of two orthogonal properties P^1 and P^2 is a simple example of dependent properties. In fact, if $\alpha_i \in R_1$ with $i \in A$, A denoting a certain continuous—or discrete—set of indices, then R_2 will contain all and only the elements $1 - \alpha_i$, $i \in A$. R_{12} is formed by all and only the pairs $(\alpha_i, 1 - \alpha_i)$ with $i \in A$ and is then strictly contained in $R_1 \times R_2$ (except the trivial case when $\#A = 1$).

Let us consider $M(M > 1)$ independent measurable properties $P^1,..., P^M$ defined in U. It is not meaningful to compare measures referring to different properties. Because of this, as we said in Section 3, a *poset* and no longer a chain is obtained as range of the generalized characteristic function defined on U. The matrix of the values

$$\begin{bmatrix} \psi_{P^{j_1}}(x) \\ \vdots \\ \psi_{P^{j_k}}(x) \end{bmatrix}$$

of the measures of the properties $P^{j_1},..., P^{j_k}$ taken into account in a certain element x of U can be interpreted as the "matrix degree" to which x enjoys the *conjunction* $P^{j_1} \wedge \cdots \wedge P^{j_k}$ of the considered properties. If in U the orthogonal properties $\sim P^{j_1},..., \sim P^{j_k}$ are also defined then by means of (5.2) one can compute, starting from the previous matrix, the degree to which x enjoys each conjunction $\gamma_1(P^{j_1}) \wedge \cdots \wedge \gamma_k(P^{j_k})$, where $\gamma_s(P^{j_s})$ stands for either P^{j_s} or $\sim P^{j_s}$. In such a case the degree matrix is

$$\begin{bmatrix} \psi_{\gamma_1(P^{j_1})}(x) \\ \vdots \\ \psi_{\gamma_k(P^{j_k})}(x) \end{bmatrix}.$$

We observe that, also in the case of a single property P, it is not always meaningful to compare two measures. In fact a measure of P can be very "rough" in some objects (for instance, ψ_P can assume only few values in [0, 1]: 1 for a "big" object, 0 for a "small" one, 1/2 in all the other cases), or very "sophisticated" for other objects, being sensitive to very small variations of the property. This lack of comparability occurs any time one makes measurements with instruments of different precisions. In this case, besides the numerical evaluation, one also gives the *precision* of the instrument used, defined as the distance between the next two lines. To express in mathematical terms the concept of "precision" of a measuring instrument, let us give the following:

DEFINITION 5.3. *For any subset R of the interval* [0, 1] *the tolerance is the map*

$$\tau_R : R \to [0, 1] \tag{5.6}$$

defined as

$$\tau_R(y) \equiv \min[| w - y |, w \neq y \text{ and } w \in R]. \tag{5.7}$$

If P is a measurable property in U having measure ψ and R_ψ = range ψ, then $\tau_{R_\psi}(y)$ with $y \in R_\psi$ gives the *minimum distance* between the degree y of the property P and any other degree. If τ_ψ is constant in U we shall call P *homogeneously measured* in U by ψ. For any ψ one can consider the class C_ψ of all measures $\{\psi_\alpha\}$ of P such that

$$| \psi_\alpha(x) - \psi(x)| < \tau_{R_\psi}(\psi(x)), \qquad x \in U.$$

If the measurements are performed by an instrument whose lines are just the values of ψ, then one is not able to discriminate with this apparatus two "measures" belonging to the class C_ψ. Therefore the introduction of real physical measuring instruments is equivalent to considering measures whose values are defined except for the tolerance τ_ψ.

6. INTERPRETATION OF THE ENTROPY OF L-FUZZY SETS

In Section 4 the notion of entropy of an L-fuzzy set was introduced in a purely formal manner, just giving a more general and extensive definition than the one considered in the case in which $L \equiv [0, 1]$. In this section a possible interpretation of the previous notion is given and its possible use in decision theory is suggested. Let us stress that in order to give an interpretation of the entropy of an L-fuzzy set as a measure of the amount of uncertainty arising in decision taking we must first clarify the kind of decision one has to consider.

Let us consider M independent completely measurable properties P defined in U, $j \in J$, $J \equiv \{1,..., M\}$. For any $P^j, j \in J$, we suppose that also the orthogonal property $\sim P^j$, $j \in J$, is defined in U.

Let $I \equiv \{x_1 ,..., x_N\}$ be a finite subset of U. We suppose that for any object of I a variable number of properties, less than or equal to M, is considered. In such a way one has an L-fuzzy set f defined on I and M maps

$$\alpha_j \circ f: A_j(f) \to [0, 1], \qquad j \in J,$$

where $\alpha_j \circ f(x_k) \equiv \psi^j(x_k)$ for $x_k \in A_j(f)$. Therefore $\alpha_j \circ f(x_k)$ has the meaning of *degree to which* x_k *enjoys the property* P^j *and* $1 - \alpha_j \circ f(x_k)$ *of degree to which* x_k *enjoys* $\sim P^j$.

We can then interpret the component $d_j(f)$ of the entropy matrix $\mathscr{d}(f)$ *as a measure of the total amount of uncertainty* arising in a decision $(P^j, \sim P^j)$ for the elements of the subset $A_j(f)$ of I.

We stress that the above interpretation is correct if the precision of the measures of the property P^j is supposed to be *homogeneous*, that is, the map τ_{ψ^j} is *equal to a constant*. Otherwise to add the different contributions of the entropies due to different objects of I can be meaningless.

The homogeneity of a *measure scale* is generally an implicit assumption in the case of a single property. However, a stronger requirement we shall first make is that all the scales are homogeneous among themselves, that is, all the maps τ_{ψ^j}, $j \in J$, are equal to a same constant. In this case let us introduce for any $x_k \in I$ the set $R[f(x_k)]$ of the indices of all considered properties. We shall denote by $D(x_k)$ a *decision in* x_k, *that is, any choice between* P^{j_s} *and* $\sim P^{j_s}$ *for any* $j_s \in R[f(x_k)]$. In such a way one can give to the quantity $\delta(f)$, formally defined by (4.6) $\delta(f) = \sum_{j=1}^{M} d_j(f)$ the meaning of *total amount of uncertainty arising for all the decisions* $D(x_k)$ *when* x_k *varies in* I.

If we do not make the hypothesis of homogeneity for the properties P^j, $j \in J$, one cannot assume $\delta(f)$ as a measure of the total uncertainty in decision taking since one cannot add the quantities $d_j(f)$, $j \in J$. In this case instead *of a scalar quantity one has to consider, as measure of uncertainty, the matrix quantity* $\mathscr{d}(f)$, *defined by* (4.4).

If f_1 and f_2 are two L-fuzzy sets such that $\mathscr{d}(f_1) \leqslant \mathscr{d}(f_2)$, one can certainly state that the first one is "sharper" than the second in the sense that, relatively to all considered properties, the total uncertainty related to f_1 is less than or equal to the one of f_2. Furthermore $\mathscr{d}(f) = 0$ if and only if there is no uncertainty in decision taking and $\mathscr{d}(f)$ is maximum, for any fixed cover of I, when $\alpha_j \circ f(x)$ are equal to $1/2$ for all the properties. We stress that in this case the entropies of L-fuzzy sets *are not always comparable*.

The meaning of the normalized entropy $\nu(f)$ defined by (4.12) is that *of average uncertainty*, or *uncertainty per decision*. If $\nu(f_1) \leqslant \nu(f_2)$, this means that the first L-fuzzy set is *on the average* sharper than the second one. The infinite value given to $\nu_j(f)$ in the case in which $A_j = \phi$ is intended to distinguish between the case in which the property P^j has not been taken into account from the one in which the function $\alpha_j \circ f$ takes the value 0 or 1 so that the entropy $d_j(f)$ is 0.

As we said in Section 4 the partial order relations between entropies and normalized entropies, except particular cases, *are not comparable*. Indeed, for instance, in the case of two properties P^1 and P^2, if for the most objects of I the property P^1 is taken into account, $d_1(f)$ can be very high and $d_2(f)$ very small, whereas the uncertainty per decision $\nu_1(f)$ can be very much smaller than $\nu_2(f)$.

We now propose to give a further interpretation of the entropy $d(f)$ of an L-fuzzy set. To this end we introduce some *composition operations* on the class of measurable properties. If P and Q are two measurable properties in U, also their disjunction $P \vee Q$ and conjunction $P \wedge Q$ are measurable with respect to the measures $\psi_{P\vee Q}$ and $\psi_{P\wedge Q}$ defined as

$$\begin{aligned} \psi_{P\vee Q}(x) &= \max\{\psi_P(x), \psi_Q(x)\}, \\ \psi_{P\wedge Q}(x) &= \min\{\psi_P(x), \psi_Q(x)\}, \qquad \text{for all} \quad x \in U. \end{aligned} \tag{6.1}$$

Let us now consider M pairs of orthogonal properties $(P^j, \sim P^j)$, $j \in J$, defined in U.

If we consider the disjunction P of the properties $P^j, j \in J$,

$$P(x) \leftrightarrow P^1(x) \vee \cdots \vee P^M(x), \qquad x \in U,$$

the measure of P according to $(6.1)_1$ is given by

$$\psi_P(x) = \max\{\psi_{P^1}(x), \ldots, \psi_{P^M}(x)\}.$$

$\psi_P(x)$ for any $x \in U$ gives the *degree to which x belongs* to the set

$$P_1(U) \equiv \bigcup_{j=1}^{M} P_1{}^j(U).$$

We observe that in U one has

$$\sim P(x) \leftrightarrow \sim P^1(x) \wedge \cdots \wedge \sim P^M(x)$$

so that the measure of $\sim P$ according to $(6.1)_2$ is

$$\psi_{\sim P}(x) = 1 - \psi_P(x), \qquad x \in U.$$

$\sim P$ is then measurable and orthogonal to P.

In the following, $\sim P$ will be supposed to be completely measurable.

If $I \equiv \{x_1 \cdots x_N\}$ is a subset of U, we can consider the measure $\alpha_P \circ \tilde{f}$ of P in I, having

$$\alpha_P \circ \tilde{f} = (\alpha_1 \circ f) \vee \cdots \vee (\alpha_M \circ f). \tag{6.2}$$

$\alpha_P \circ \tilde{f}$ is a fuzzy set defined in I, and for any $x_k \in I$, $\alpha_P \circ f(x_k)$ is interpreted as *the degree to which* x_k enjoys P (or belongs to $P_1(U)$) and $1 - (\alpha_P \circ \tilde{f})(x_i)$ as *the degree to which* x_k *enjoys* $\sim P$.

We stress that the value of $\alpha_P \circ \tilde{f}$ refer to different properties which are generally characterized by different *precision measures* τ_ψ.

Let first suppose all the measures to be homogeneous. In this case the *entropy* of the fuzzy set $\alpha_P \circ \tilde{f}$ *can be assumed as measuring the total amount of uncertainty in the decision taking* $(P, \sim P)$ *for all the objects of* I.

One can consider $\tilde{f}$ as an L-fuzzy set formed by M chains if for any $x_k \in I$ there exists only an index j such that (6.2) is satisfied, that is,

$$\alpha_P \circ \tilde{f}(x_k) = \alpha_j \circ f(x_k).$$

In this case one has

$$d(\alpha_P \circ \tilde{f}) = \delta(\tilde{f}) = \sum_j d_j(\tilde{f}).$$

If there exist more indices j such that satisfy (6.2), denoting this set by $W[\tilde{f}(x_i)]$, one can measure the uncertainty by the quantity:

$$u(\alpha_P \circ \tilde{f}) = \sum_i \sum_j \frac{1}{\#W[\tilde{f}(x_i)]} T(\alpha_j \circ \tilde{f}(x_i)). \tag{6.3}$$

Let us now consider the case of *not homogeneous* measures. In this case if $\#W[\tilde{f}(x_k)] = 1$ for all $x_k \in I$ the matrix $\mathscr{d}(\tilde{f})$ can be assumed as a measure of the total uncertainty in the decision taking $(P, \sim P)$ *for all the elements of* I. $\mathscr{d}(\tilde{f})$ is 0 if and only if all the elements of I either enjoy P or $\sim P$; $\mathscr{d}(\tilde{f})$ is *maximum*, for a fixed partition, if and only if $\alpha_P \circ \tilde{f}(x)$ holds $1/2$, for all the elements of I.

We stress, however, that the entropies $\mathscr{d}$ are not always comparable. If there are more indices j such that (6.2) is satisfied a (matrix) measure in the decision taking $(P, \sim P)$ can be given by the matrix $\mathscr{u}(\tilde{f})$ whose components are

$$u_j(\tilde{f}) = \sum_i \frac{1}{\#W[\tilde{f}(x_k)]} T(\alpha_j \circ \tilde{f}(x_k)).$$

In the case of homogeneity of measures one has

$$u(\alpha_P \circ \tilde{f}) = \sum u_j(\tilde{f}).$$

We observe that the matrix $\mathscr{U}(f)$ can be expressed as

$$\mathscr{U}(f) = \frac{1}{\prod_k \# W[\tilde{f}(x_k)]} \sum_\alpha \mathscr{d}_\alpha(\tilde{f}), \tag{6.4}$$

where $\{\mathscr{d}_\alpha(\tilde{f})\}$ is the set of all entropy matrices that one can obtain disposing, for any $x_k \in I$, $\tilde{f}(x_k)$ in any one of the chains of order index belonging to $W[\tilde{f}(x_k)]$.

Let us stress that the above rule of giving the some contribution

$$\frac{1}{\# W[\tilde{f}(x_k)]} T(\alpha_j \circ \tilde{f}(x_k))$$

to all chains whose index belongs to $W[\tilde{f}(x_k)]$, in the construction of the uncertainty matrix $\mathscr{U}(f)$, leads to the result (6.4), that is quite natural to expect. In the case of a possible set of entropy matrices, in absence of other information, one can consider as a measure of uncertainty the *ensemble average* of these matrices; that is equivalent to consider them on the same footing for what concerns the information they give.

7. Concluding Remarks

Shannon's entropy plays an essential role in information theory. Apart from the similarity with the thermodynamic entropy, the main argument for the use of this quantity is the fact that the basic *coding theorem* of information theory is true only for Shannon's entropy.

However, the entropy measures can be used not only in problems of *transmission of information*, but also in other fields. For this reason in some applications other entropy measures have been introduced (see, for instance, Schützenberger, 1953; Rényi, 1960).

It has been shown (Vajda, 1969) that in *statistical pattern recognition a class* of entropy measures (see Eq. 2.4) which includes Shannon's entropy can be taken into account. These measures enjoy only some of the properties of Shannon's entropy as the basic one of the concavity of the function $-p \ln p$.

Often the use of entropies different from Shannon's is more convenient in order to relate the uncertainty measures to some typical quantities of the decision-taking processes. The functional one has to use generally depends on the particular considered problems.

One can also consider measures of uncertainty outside of a statistical context, in decision-taking processes performed on ensembles of objects described by fuzzy sets (De Luca and Termini, 1972; Capocelli and De Luca, 1972).

It has been shown (Capocelli and De Luca, 1972) that some measures of entropy and "energy" of the considered systems can be easily related to the results of *deterministic or probabilistic* decisions performed on them.

A further and more substantial generalization of the classical concept of entropy has been presented in this paper. The entropy is no longer a *numerical quantity* but a *column matrix* (or vector). This kind of quantity is considered for a macroscopic description of systems formed by objects individually described by a given set of *independent* and *measurable properties.*

In such a way not all the possible situations (or *macroscopic states*) may be compared by means of their entropies.

It seems to us that the previous situation occurs very frequently in pattern recognition where often the systems are described by means of independent measurable properties that may have different weights in order to classify the patterns, and whose measures scales only in some cases are homogeneous.

As we have seen in Section 5 an interpretation of the previous entropy is attained in decision processes: its meaning is still the one of an uncertainty (matrix) measure.

We note, however, that some specifications of the decisions process and further information on the homogeneity and relevance of the properties may allow extracting from the matrix-entropy a numerical evaluation of the uncertainty.

We think that the previous generalization of the classical concept of entropy are necessary in order to attain a "good" macroscopic description of the systems on which decision processes take place.

RECEIVED: April 19, 1973

REFERENCES

CAPOCELLI, R. M., AND DE LUCA, A. (1973), Fuzzy sets and decision theory, *Information and Control* **23**, 446–473.

CAPOCELLI, R. M., AND DE LUCA, A. (1972), Measures of uncertainty in the context of fuzzy sets theory, 2° *Congresso Nazionale di Cibernetica, Casciana Terme*, pp. 235–242.

DE LUCA, A., AND TERMINI, S. (1970), Algebraic properties of fuzzy sets, *Notices Amer. Math. Soc.* **17**, 944 abstract; *J. Math. Anal. Appl.* **40**, 373–396.

DE LUCA, A., AND TERMINI, S. (1972), A definition of a non-probabilistic entropy in the setting of fuzzy sets, *Information and Control* **20**, 301–312.

GOGUEN, J. A. (1967), *L*-Fuzzy sets, *J. Math. Anal. Appl.* **18**, 145–174.

KOOPMAN, B. O. (1940), The bases of probability, *Bull. Amer. Math. Soc.* **46**, 763–774.
ŁUKASIEWICZ, J., AND TARSKY, A. (1930), Untersuchungen über den Aussagenkalkul, *C. R. Soc. Sci. Lett. Varsovie III* **23**, 1–21, 30–50.
RÉNYI, A. (1960), On measures of entropy and information, *Proc. Berkeley Symp. Math. Statist. and Probability, 4th, 1961* **1**, 547–561.
SCHÜTZENBERGER, M. P. (1954), Contribution aux applications statistiques de la théorie de l'Information, *Publ. Inst. Statist. Univ. Paris* **3**, 3–117.
VAJDA, I. (1969), A contribution to the informational analysis of pattern, "Methodologies of Pattern Recognition" (S. Watanabe, ed.), pp. 509–519, Academic Press, New York.
ZADEH, L. A. (1965), Fuzzy sets, *Information and Control* **8**, 338–353.