INFORMATION AND CONTROL 23, 446–473 (1973)

Fuzzy Sets and Decision Theory

R. M. CAPOCELLI AND A. DE LUCA

Laboratorio di Cibernetica del C.N.R., Via Toiano, 2 (80072) Arco Felice, Naples, Italy

The problem of making decisions to classify the objects of a certain universe into two or more suitable classes has been considered in the setting of *fuzzy sets* theory. A measure of the total amount of uncertainty that arises in making decisions has been proposed in the general case. This quantity reduces to the "entropy" of a fuzzy set in the case of two classes. Other quantities which play a relevant role in this theory are the "energy" and the "effective power" of a fuzzy set, defined as

$$\sum_{i=1}^{N} w_i f_i \quad \text{and} \quad \phi \sum_{i=1}^{N} f_i ,$$

respectively, where w is a nonnegative weight function and ϕ a nonnegative constant. If w = constant and $\phi \neq 0$, the energy is proportional to the effective power and, therefore, to the "power" of the fuzzy set. The maximum of the uncertainty has been calculated in some cases of interest, keeping constant the total energy and effective power. In particular the Maxwell–Boltzmann, Fermi–Dirac, and Bose–Einstein distributions are derived. Some applications to decision theory are considered in the case of both deterministic and probabilistic decisions. Finally, the analogies that exist between the previous concepts and the thermodynamic ones are discussed.

1. INTRODUCTION

The theory of "making decisions," that plays a fundamental role in many scientific branches, has been mainly developed in the setting of probability theory. Although probabilistic decision-methods (see, for instance, Wald (1971)) often work very well in many fields, such as *pattern recognition*, there exist cases in which these methods are *ineffective*. This occurs whenever the standard probabilistic formalism is not appropriate for the description of the considered situations; for instance, when the latter are not really *random* so that the introduction of probabilities as measures of empirical frequencies in a large number of identical experiments may become meaning-less. Often the "source" of uncertainty that arises in decision-making may be in part or even completely *deterministic* (De Luca and Termini, 1972).

In this paper we shall provide a general analysis of such deterministic sources of uncertainty, in the setting of "fuzzy sets" theory. As we will try to show, this theory introduced by Zadeh (1965) is of great importance in *decision theory* where, as we said above, sometimes the usual concepts and methods of *probability theory* either are not appropriate or do not yield relevant results.

In a previous paper (De Luca and Termini, 1972) a global measure of the "degree of fuzziness" of a fuzzy set has been proposed. This quantity has been called "entropy" of a fuzzy set even though its meaning is quite different from the one of Shannon's *information theory* (Shannon and Weaver, 1962). Indeed no random experiments are needed in order to define it. This entropy can also be interpreted as measuring the total "amount of uncertainty" that arises if, considering a fuzzy set defined on a class of objects, one has to make decisions in order to attribute to them the "presence" or "absence" of a certain property. It is also possible to associate to this entropy a "quantity of information" received when the decisions have been taken, so that the previous uncertainty vanishes.

This approach to fuzzy sets theory has been extended providing a *measure of uncertainty* in the case of more properties, which satisfy a condition called of "orthogonality" (see also Capocelli and De Luca, 1972b). Moreover, other "thermodynamic" concepts such as the "energy" or the "effective power" of fuzzy sets are introduced. The maximum of the uncertainty has been calculated in some cases of interest keeping constant the total energy and the effective power of the fuzzy sets thus obtaining, in particular, the *Maxwell–Boltzmann* distribution and the *intermediate statistics*, particular cases of which are the *Fermi–Dirac* and the *Bose–Einstein* distributions. Even though these distributions are derived in a way similar to the one followed in *statistical-thermodynamics*, the interpretation of the quantities involved differs from the particle-physical one and, generally, does not require a *probabilistic–frequentistic* context.

It seems to us that by means of the concepts of "energy" and "power," it is possible to give a thermodynamic-like treatment of the "fuzziness" which has not been done so far for the "information process" in the setting of Shannon's information theory, since a concept equivalent to thermodynamic energy has not been introduced.

Some applications to decision theory are considered in the case of both *deterministic* and *probabilistic* decisions. In the latter the probability distribution of the energy of the classical set obtained has an *average value* equal to the *energy of* the fuzzy set and an upper bound for the *variance* proportional to the *entropy*. In conclusion the analogies that exist between our approach to fuzzy sets theory and *thermodynamics* are briefly discussed.

2. Measurable Properties

One of the most typical "sources" of uncertainty in the classification of the objects of a certain universe into two or more classes occurs when the objects may enjoy to a different degree the properties which characterize the classes themselves. In this section we introduce a mathematical formalism that enables us to describe this kind of uncertainty which occurs very frequently in decision theory and, generally, is not related to random experiments. Indeed, here we are not considering objects which may belong to the given classes, and only to them, with certain probabilities, but objects which belong to either or none of the given classes (for instance, the grey objects which are neither *white* nor *black*). Of course, this uncertainty may be reduced or eliminated if one changes the language of the description by including in it new concepts or classes. As stressed in De Luca and Termini (1971), the previous situation very often occurs in making physical and cybernetical models: If we try to describe new situations in terms of some classical concepts one has to use a language that yields an unavoidable lack of information about the considered system.

Let a set U be given and a property P defined in it. We denote by $P_1(U)$ and $P_{\neq 1}(U)$ the subsets of U formed by the elements for which P is *true* and *false*, respectively. Let further $P_0(U)$ be a subset of $P_{\neq 1}(U)$; we can then consider the property $\sim P$ defined in U as

$$\sim P(x) \leftrightarrow x \in P_0(U). \tag{2.1}$$

If $\neg P$ denotes the *complementary property* of P, in U there results

$$\sim P(x) \rightarrow \neg P(x); \tag{2.2}$$

whereas, the *inverse implication* is true only in the subset of U given by $A(U) \equiv P_1(U) \cup P_0(U)$.

In this way by any subset $P_0(U)$ of $P_{\neq 1}(U)$ we can obtain a further property $\sim P$ such that if $\sim P$ is true P is false but the converse holds only in the subset $A(U)$ of U. We shall now see how to specify a subset $P_0(U)$ of $P_{\neq 1}(U)$. To this purpose, we assume that the elements of U can enjoy, to a different degree, the property P. Furthermore, we suppose we shall be able to provide, for each element x of U, a measure of how much x enjoys the property P. Formally we can give a map

$$\psi_P: U \rightarrow [0, 1] \tag{2.3}$$

from U to the interval [0, 1] of the real line such that

$$P(x) \leftrightarrow (\psi_P(x) = 1). \tag{2.4}$$

For any x, $\psi_P(x)$ is called a *measure* of the property P in x and interpreted as *degree to which x enjoys the property P*, or *degree to which x belongs to the set $P_1(U)$*. For this reason ψ_P is also named *membership function*. If $\psi_P(x) = 1$ we say that the property P is "present" in x. If $\psi_P(x) = 0$ we say that the property P is "absent" in x. In the following we shall identify $P_0(U)$ with the subset of $P_{\neq 1}(U)$ formed by the elements of U in which the property P is absent. A property P defined in U, such that the previous assumptions are satisfied will be called "measurable" in U. If, in addition, we suppose that $P_1(U)$ is not *empty*, P will be said to be "completely measurable."

In the following we shall confine ourselves to considering only the case when U is such that its elements enjoying neither P nor $\sim P$ have features "intermediate" between those of $P_1(U)$ and $P_0(U)$. We can mathematically express this circumstance by supposing that P and $\sim P$ are completely measurable and such that

$$\psi_{\sim P}(x) = 1 - \psi_P(x), \qquad \text{for all } x \in U. \tag{2.5}$$

Two measurable properties satisfying (2.5) will be named "orthogonal," rather than complementary. Indeed they are complementary, from a lattice-theoretic point of view, only in the subset $A(U)$ of U; whereas, in $A - A(U)$, supposed to be nonempty, the condition $P(x) \vee \sim P(x) = T$ ($\vee$ denoting the disjunction and T the property always true) is not satisfied but replaced by the condition on the measures (2.5). We can easily see that in U, $\sim(\sim P)$ is equivalent to P and that $\sim(\neg P(x)) \rightarrow \neg(\sim P(x))$ so that the two operations $\neg$ and $\sim$ *do not commute* except when x belongs to $A(U)$, where $\neg$ and $\sim$ are equivalent.

Let P and $\sim P$ be measurable in U; while we can decide, for any element x of U, whether x enjoys either P or $\neg P$, this is possible for P and $\sim P$ only in the subset $A(U)$ of U. This fact arises whenever, starting from a given universe $A(U) \equiv P_1(U) \cup (\sim P)_1(U)$ on which two complementary properties P and $\sim P$ are defined, we enlarge $A(U)$ by including other objects which enjoy neither P nor $\sim P$. A noteworthy case is when P and $\sim P$

are completely measurable and the condition (2.5) holds: That is the added objects have intermediate features between those enjoying P and $\sim P$. In such a situation if we have to decide whether to attribute an object x to $P_1(U)$ or to $P_0(U)$, an "ambiguity" or "uncertainty" arises which is *maximum* when $\psi_P(x)$ takes the value $1/2$ and decreases when $\psi_P(x)$ goes to 0 or 1; in these two cases the ambiguity vanishes and $x \in A(U)$.

We remark that generally one has to "measure" the presence of a certain property P (that is, to evaluate the map ψ_P) only over finite subsets of a given universe U. If $I \equiv \{x_1, \ldots, x_N\}$ is such a subset, let us denote by f the *restriction* of ψ_P to I, that is the map

$$f: I \rightarrow [0, 1]$$

such that $f(x_k) = \psi_P(x_k)$ $(k = 1, \ldots, N)$. We shall call f *measure of P in I*. In general it is difficult to give an interpretation to f in I without taking U into account. For instance, $P_1(I)$ or $P_0(I)$ or both could be empty. Moreover, the property $P(I)$ defined in U as $P(I, x) \leftrightarrow x \in P_1(I)$ might determine in U a completely measurable subproperty of P such that the degree to which some elements of I enjoy $P(I)$ is different with respect to the degree to which they enjoy P. The same holds for $\sim P(I)$ defined in U as $\sim P(I, x) \leftrightarrow x \in P_0(I)$.

However, it is possible to interpret $f(x_i)$ $(i = 1, \ldots, N)$ as the *degree to which x_i belongs to $P_1(I)$* (or enjoys $P(I)$) if the measure $\psi_{P(I)}$ of the completely measurable property $P(I)$ restricted in I coincides with f, that is

$$\psi_{P(I)}(x_i) = \psi_P(x_i) = f(x_i), \qquad (i = 1, \ldots, N).$$

Furthermore, to speak of ambiguity in a decision $P(I)$, $\sim P(I)$ requires the validity in I of the condition (2.5), which becomes

$$\psi_{\sim P(I)}(x_i) = 1 - \psi_P(x_i) = 1 - f(x_i), \qquad (i = 1, \ldots, N).$$

The extension of the previous considerations to the case of M properties is straightforward. M measurable properties $P^0, \ldots, P^{M-1}$ on U will be called *orthogonal* if their measure functions $\psi^0, \ldots, \psi^{M-1}$ satisfy the conditions

$$\sum_{j=0}^{M-1} \psi^j(x) = 1, \qquad \text{for all } x \in U. \tag{2.6}$$

Moreover, if the properties P^j $(j = 0, \ldots, M-1)$ are completely measurable and we have to attribute an object x to $P_1^0(U)$ or $P_1^1(U), \ldots,$ or to $P_1^{M-1}(U)$ an uncertainty arises that is maximum when $\psi^j(x) = 1/M$ $(j = 0, \ldots, M-1)$ and is 0 on the set $A(U) \equiv \bigcup_{k=0}^{M-1} P_1^k(U)$. In such a case, from (2.6), only one of $\psi^j(x)$ holds 1 all the others being 0.

To conclude this section, we stress that in order to introduce measures of uncertainty in decision-making, that will be the subject of the next section, we do not need further assumptions on the property measures ψ^j ($j = 0,..., M - 1$). However, we note that in many cases it is meaningful to compose the measurable properties in the following way: If P and Q are two measurable properties, so are the disjunction $P \vee Q$ and the conjunction $P \wedge Q$ with respect to the measures $\psi_{P\vee Q}$ and $\psi_{P\wedge Q}$ defined as

$$\begin{aligned} \psi_{P\vee Q}(x) &\equiv \max\{\psi_P(x), \psi_Q(x)\}, \\ \psi_{P\wedge Q}(x) &\equiv \min\{\psi_P(x), \psi_Q(x)\}, \qquad \text{for all } x \in U. \end{aligned} \tag{2.7}$$

The previous operations have been used by Zadeh (1965) to define *union* and *intersection* of fuzzy sets.[1] From $(2.7)_1$ we see that, in place of the *addition* rule for probabilities, one has that the measure of a disjunction equals the *largest* of measures of components. In such a way it follows that, apart from interpretative problems, there are substantial differences between the previous formalism and the one of probability theory.

3. Measure of Uncertainty

Let us consider a finite subset I of a universal class U on which a certain number of completely measurable and orthogonal properties are defined. Our aim is to give a *measure of the total amount of uncertainty* that arises in deciding for every object of I which of the considered properties is enjoyed. We make the assumption that the next decisions are "independent" in the sense that any decision does not yield changes in the uncertainty present before the others. In the case of two properties a measure of the total uncertainty has been given in De Luca and Termini (1972) by means of the concept of *entropy of a fuzzy set.*

Subsequently, a generalization and extension of the previous measure to the case of more than two orthogonal properties has been proposed (Capocelli and De Luca, 1972b). Along this line, we shall consider here, in a more general and rigorous way, the problem of the measure of the uncertainty in decision-making in the setting of fuzzy sets theory.

Let P^j ($j = 0,..., M - 1$) be M completely measurable properties defined in U and f^j ($j = 0,..., M - 1$) their measures in I, which are *fuzzy sets*

[1] An algebraic analysis of fuzzy sets which respect the previous operations and the relationships existing with probability theory can be found in De Luca and Termini (1970).

defined in I. We shall denote by f the M-tuple $f^0,\dots,f^{M-1}$ and by $\mathscr{L}(I)$ the class of *all* fuzzy sets definable in I. We suppose that the properties P^j ($j = 0,\dots, M-1$) are *orthogonal*, so that for any element x_i of I, the fuzzy sets f^j ($j = 0,\dots, M-1$) satisfy the condition,

$$\sum_{j=0}^{M-1} f^j(x_i) = 1, \qquad (i = 1,\dots, N). \tag{3.1}$$

For any integer M, we now introduce the following functional

$$u(f) \equiv u(f^0,\dots,f^{M-1}) \equiv \sum_{i=1}^{N} u(f_i). \tag{3.2}$$

N is the number of elements of I and $u(f_i)$ is the function

$$u(f_i) \equiv u(f_i^0,\dots,f_i^{M-1}) \equiv \sum_{j=0}^{M-1} v(f_i^j), \tag{3.3}$$

where f_i^j stands for $f^j(x_i)$ and v is a *continuous strictly concave* function[2] (Hardy, Littlewood and Pólya, 1967) in the interval $(0, 1)$ such that $v(1) = v(0) = 0$. It is also possible to express v as $v(x) = xL(1/x)$, L being a continuous concave function in $[1, +\infty)$ such that $L(1) = 0$ and

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

In such a way (3.3) becomes

$$u(f_i) = \sum_{j=0}^{M-1} f_i^j L(1/f_i^j). \tag{3.4}$$

L functions that satisfy the previous assumptions are, for instance, $L_1(x) = \ln x$ and $L_2(x) = 1 - 1/x$.

In these cases (3.4) becomes

$$u_1(f_i) = -\sum_{j=0}^{M-1} f_i^j \ln f_i^j \tag{3.5}$$

[2] A measure like as (3.3), with v continuous and concave in $(0, 1)$, has been recently considered by some authors within the framework of a statistical approach to pattern recognition problem (see, e.g., Vajda, 1969).

and

$$u_2(f_i) = \sum_{j=0}^{M-1} f_i^j(1 - f_i^j), \tag{3.6}$$

respectively.

In the case $M = 2$, (3.5) reduces to $S(f_i)$ where $S(x)$ is Shannon's function $-x \ln x - (1 - x) \ln(1 - x)$, and $u(f)$ to the *entropy of a fuzzy set* (De Luca and Termini, 1972)

$$d(f) \equiv \sum_{i=1}^{N} S(f_i) = -\sum_{i=1}^{N} [f_i \ln f_i + (1 - f_i) \ln(1 - f_i)]. \tag{3.7}$$

Let us now assume the universe U to be such to include, for all the properties P^j $(j = 0,..., M - 1)$, elements which enjoy them to all possible degrees, such that (3.1) is verified; $u(f_i)$, for every i, satisfies the property.

P_1—*for any* i, *one has* $0 \leqslant u(f_i) \leqslant L(M)$. *The value* 0 *is taken if, and only if, all* f_i^j $(j = 0,..., M - 1)$ *equal* 0, *except one which equals* 1; *the value* $L(M)$ *is taken if, and only if,* $f_i^j = 1/M$ $(j = 0,..., M - 1)$.

We assume $u(f_i)$ as a *measure* of the uncertainty that arises in deciding which of properties P^j $(j = 0,..., M - 1)$ is enjoyed by the object x_i, and $u(f)$ as a measure of the *total amount of uncertainty* for all the decisions, in the hypothesis of independent decisions.

Apart from the property P_1, which insures that the maximum of the uncertainty $u(f_i)$ is reached when all the degrees f_i^j $(j = 0,..., M - 1)$ are equal, in order for $u(f)$ to be a "good" measure of uncertainty, one must verify that it decreases for any intuitively less ambiguous situation, such as the one obtained by increasing the maximum of the f_i^j $(j = 0,..., M - 1)$ and reducing all the others.

We can prove that the following property holds.

P_2—*If* y^i $(i = 0,..., M - 1)$ *and* $\hat{y}^i$ $(i = 0,..., M - 1)$ *are two sets of nonnegative numbers such that*

$$\sum_{i=0}^{M-1} y^i = \sum_{i=0}^{M-1} \hat{y}^i$$

and $y^0 \leqslant y^1 \cdots \leqslant y^{M-1}, \hat{y}^i \leqslant y^i$ $(i = 0,..., k), \hat{y}^j \geqslant y^j$ $(j = k + 1,..., M - 1)$, *then* $u(\hat{y}) \leqslant u(y)$.

Indeed, the function v, being a continuous concave function in (0.1), has everywhere left and right derivatives v_l' and v_r' both nonincreasing and such that $v_l'(x) \geqslant v_r'(x)$, for all $x \in (0, 1)$ (Hardy, Littlewood, and Pólya, 1967, Chapter 3, pp. 91–94). We can then derive that

$$u(\hat{y}) - u(y) = \sum_{i=0}^{k} [v(\hat{y}^i) - v(y^i)] + \sum_{i=k+1}^{M-1} [v(\hat{y}^i) - v(y^i)]$$

$$\leqslant v_l'(y_k) \sum_{i=0}^{k} (\hat{y}^i - y^i) + v_r'(y_k) \sum_{i=k+1}^{M-1} (\hat{y}^i - y^i)$$

$$\leqslant v_l'(y_k) \left[\sum_{i=0}^{M-1} \hat{y}^i - \sum_{i=0}^{M-1} y^i \right] = 0.$$

In particular, if $y^i = 1/M$ $(i = 0,..., M-1)$ and $\hat{y}^0 = 0$, $\hat{y}^i = 1/M - 1$ $(i = 1,..., M-1)$, it follows that $L(M-1) \leqslant L(M)$ so that $L(M)$ is a nondecreasing function of M.

From (3.2) and P_1 we have that, for all M and N,

$$0 \leqslant u(f) \leqslant NL(M).$$

We now wish to point out some further properties of the uncertainty $u(f)$ that follow directly from definition (3.2).

P_3—*$u(f) = u(Pf)$ when Pf denotes any permutation of the M-tuple $(f^0,..., f^{M-1})$. Moreover $u(f) = u(Qf)$ where Qf denotes any M-tuple obtained from $(f^0,..., f^{M-1})$ by permuting the elements of I.*

P_4—*Denoting by $P(f^j)$ the power $\sum_{i=1}^{N} f_i^j$ of the fuzzy set f^j* (De Luca and Termini, 1972), *one has from concavity of the function $xL(1/x)$*

$$u(f) \leqslant \sum_{j=0}^{M-1} P(f^j) L\left[\frac{N}{P(f^j)}\right]. \tag{3.8}$$

P_5—*Denoting by $h(f^j)$ $(j = 0,..., M-1)$ the quantity*

$$h(f^j) \equiv \sum_{i=1}^{N} T(f_i^j), \tag{3.9}$$

having set

$$T(x) \equiv xL(1/x) + (1-x) L(1/1-x), \tag{3.10}$$

one has

$$u(f) \leqslant \sum_{j=0}^{M-1} h(f^j), \tag{3.11}$$

where the equality holds if and only if $u(f) = 0$.

If A and I are two subsets of U one can consider the uncertainties $u(A, f_A)$ and $u(I, f_I)$, where $f_A \equiv (f_A{}^0, \ldots, f_A^{M-1})$ and $f_I \equiv (f_I{}^0, \ldots, f_I^{M-1})$; $f_A{}^j$ and $f_B{}^j$ ($j = 0, \ldots, M-1$) denote the restriction of ψ^j to A and B, respectively. If $A \subseteq I$, from the definition (3.2) we get

$$u(I, f_I) = u(A, f_A) + u(I - A, f_{I-A}). \tag{3.12}$$

In particular we have

$$u(I, \hat{f}_I) = u(A, f_A),$$

where $\hat{f}_I$ is a M-tuple of fuzzy sets defined in I such that $\hat{f}_I \equiv f_I$ in A and *all* the $\hat{f}^j$ vanish in $I - A$.

Let us now make the following strong requirement on the total uncertainty $u(f)$.

R_1—*for all the integers M and N the maximum of the total uncertainty in making N decisions on M orthogonal properties is equal to the maximum uncertainty in making one decision on M^N orthogonal properties.*

$L(M^N)$ must then equal $NL(M)$; in such a case it is easy to prove (cf. for instance, Khinchin (1957), pp. 10–12) that $L(x)$ must be equal to $\alpha \ln x$, where α is a positive constant.

Many mathematical properties of the uncertainty $u(f)$ that we shall analyze in the following and many aspects of the decision theory that is possible to develop, do not depend on this requirement, even though the use of the logarithm function often simplifies the mathematical formalism; in the following we shall always refer to the general expression (3.4) of the uncertainty, except in Sections 5 and 6.

We conclude the section emphasizing that even if we assume $L(x) = \ln x$, the uncertainty $u(f)$ we have introduced has a completely different meaning from Shannon's entropy since no random experiments are needed in order to define it. Indeed, $u(f)$ is a measure of the total amount of uncertainty in making decisions on the objects of I; whereas, Shannon's entropy gives a measure of the (statistical) average uncertainty in foreseeing the event that will occur in a random experiment.

4. Entropy, Energy, and Effective Power of a Fuzzy Set

In this section we shall introduce some macroscopic quantities associated to the fuzzy sets such as "entropy," "energy," and "effective power" which allow us to interpret some properties of the uncertainty, introduced in the previous section, in a thermodynamic-like fashion.

This analogy with thermodynamics will appear more clear in Section 5 where, in a context of decision theory, some classical and quantum distributions are derived. Moreover, in Section 6 we shall see how the previous quantities play a relevant role in decision theory.

Let f be a fuzzy set defined on a finite set I; we give the following definitions.

DEFINITION 4.1. The entropy of f is the nonnegative functional $h(f)$ defined by (3.9),

$$h(f) \equiv \sum_{i=1}^{N} T(f_i), \tag{4.1}$$

where T is given by (3.10).

From this definition it follows that the entropy of a fuzzy set is taken equal to the total uncertainty in making decisions in the case of two orthogonal properties P and $\sim P$. If L is equal to the logarithm function, $h(f)$ reduces to the entropy $d(f)$ given by (3.7). In the following, to avoid misleadings, we shall call $d(f)$ *logarithmic entropy* of f.

Often it is more meaningful to refer, instead of to the entropy $h(f)$, to the *normalized entropy* $\nu(f)$ defined as

$$\nu(f) \equiv \frac{1}{N} h(f), \tag{4.2}$$

which gives a measure of the *average uncertainty* or *uncertainty per decision.*

DEFINITION 4.2. If w is a nonnegative weight function defined on I, the energy $E(w, f)$ of the fuzzy set f is the quantity (Capocelli and De Luca, 1972a)

$$E(w, f) \equiv \sum_{i=1}^{N} w_i f_i\,. \tag{4.3}$$

DEFINITION 4.3. If ϕ is a nonnegative constant the effective power of f is the quantity

$$P(\phi, f) \equiv \phi P(f), \tag{4.4}$$

where $P(f) = \sum_{i=1}^{N} f_i$ is the power of f.

If w is a constant then the energy is proportional to the power and, if $\phi \neq 0$, to the effective power.

In the following we shall call $w_i f_i$, f_i, and ϕf_i ($i = 1, ..., N$) the *energy*, the *power*, and the *effective power* of the ith element, respectively.

As discussed in De Luca and Termini (1972) the power of a fuzzy set is an extension to fuzzy sets of the concept of *number of elements* of an ordinary set. The energy and the effective power of a fuzzy set are more general and useful concepts when the elements of I or the powers of some fuzzy sets can be differently weighted. As an example, let us consider a bundle of N wires each of which may or may not carry a signal of a fixed voltage amplitude. To each wire one can associate a variable x with values 1 or 0 according to whether the wire is active or inactive.

The quantity $\sum_{i=1}^{N} x_i$ gives the *total number* of active wires; it coincides with the total amplitude of the signals carried by the bundle. If the wires, as in a threshold elements, go through a weight-system which changes the amplitude x_i of the ith wire into $w_i x_i$, the total amplitude, after the weight-system is $\sum_{i=1}^{N} w_i x_i$. Let us now suppose that there exist some "disturbances" or malfunctions in the wires so that the amplitudes of the signals carried by the wires are between 1 and 0. In this case the bundle can be described by a fuzzy set and the total amplitude carried by the bundle coincides with the power $P(f)$ and the energy $E(w, f)$ before and after the weight-system.

Let us now suppose that instead of weighting the individual wires of a bundle, one weights the output lines $0, 1, ..., M-1$ of a set of bundles, carrying the total amplitudes of each bundle measured by $P(f^0), P(f^1), ..., P(f^{M-1})$. If $\phi^0, ..., \phi^{M-1}$ are the weights, the total amplitude of the set of bundles after the weight-system will be given by

$$\sum_{j=0}^{M-1} \phi^j P(f^j) = \sum_{j=0}^{M-1} P(\phi^j, f^j),$$

that is it equals the sum of the effective powers $P(\phi^j, f^j)$ ($j = 0, ..., M-1$).

In this section we confine ourselves to the case when the weight function w of the energy of a fuzzy set is a constant w, so that the energy of a fuzzy set is measured by the effective power $wP(f)$. From the property P_4 of the previous section follows this noteworthy proposition.

PROPOSITION 4.1. *The maximum of the uncertainty* $u(f)$ *keeping the powers* $P(f^j)$ ($j = 0, ..., M-1$) (*or the energies*) *equal to constants* P^j ($j = 0, ..., M-1$),

$$\sum_{j=0}^{M-1} P^j = N,$$

is reached when

$$f_i^j = \frac{P^j}{N} = \frac{E^j}{w^j N} \qquad (i = 1,..., N; j = 0,..., M-1),$$

that is when the energies $E^j = w^j P^j$ $(j = 0,..., M-1)$ *are equidistributed over the N objects of I.*

Denoting by $u_{\max}(E^0,..., E^{M-1})$ the maximum of the uncertainty $u(f)$ in the case of constant energies, the maximum of $u_{\max}$ for all possible energies is reached for $E^j = Nw^j/M$ $(j = 0,..., M-1)$ or, equivalently, for $f_i^j = 1/M$ $(i = 1,..., N; j = 0,..., M-1)$ so that it coincides with the maximum of $u(f)$.

The results previously obtained can be widely generalized taking into account more complex situations like, for instance, those in which the energies are constant only for *some* fuzzy sets on all or in some part of I. As an example, if the energies of k fuzzy sets f^j $(j = 0,..., k-1)$ are constant then the maximum of $u(f)$ is reached for

$$f_i^j = \frac{E^j}{Nw^j} \qquad (i = 1,..., N; j = 0,..., k-1),$$

$$f_i^j = \frac{1 - \sum_{s=0}^{M-1} E^s/Nw^s}{M-1-k} \qquad (i = 1,..., N; j = k,..., M-1).$$

Furthermore, if we fix to be a constant *not* the total energy of a fuzzy set but only the energy of a subset of I formed by k elements $(k \leqslant N)$ then under the condition of maximum uncertainty the previous amount of energy is equidistributed over the considered subset.

From the property P_5 of the previous section it follows that *the total uncertainty* $u(f) = u(f^0 \cdots f^{M-1})$ *is less or equal to the sum of the entropies of the fuzzy sets* f^j $(j = 0,..., M-1)$.

At the end of the section we introduce some other quantities associated to a fuzzy set that we call "moments" of a fuzzy set whose interpretation will be made clearer in the next sections.

DEFINITION 4.4. For any positive integer h the moment of h-order of the fuzzy set f is the quantity

$$M_h(f) \equiv \sum_{i=1}^{N} \frac{f_i^h(1-f_i) + (1-f_i)^h f_i}{2} \qquad (h = 1, 2,...). \tag{4.5}$$

From this definition it follows that the entropy $h(f)$ is zero, that is f is classical, if and only if all moments vanish. Finally we note that the first order moment, that we shall also denote by $m(f)$:

$$m(f) \equiv \sum_{i=1}^{N} f_i(1 - f_i), \tag{4.6}$$

can itself be assumed as a measure of the entropy of a fuzzy set.

5. A Particular Case

As we said in the previous section, in the case of only two orthogonal properties, the uncertainty $u(f)$ reduces to the entropy of the fuzzy set f

$$h(f) = \sum_{i=1}^{N} T(f_i). \tag{5.1}$$

The maximum of $h(f)$ keeping the power $P(f)$ (or the energy in the case w = constant) equal to a constant P, is given by $NT(P/N)$ which is a monotonically increasing function of P in the interval $[0, N/2]$, monotonically decreasing in $[N/2, N]$ with a maximum at $P = N/2$.

Identifying $T(x)$ with Shannon's function, (5.1) reduces to the logarithmic entropy $d(f)$ of a fuzzy set. In this case the maximum $D(N, P)$ of the entropy is given by

$$D(N, P) \equiv -P \ln P/N - (N - P) \ln (N - P)/N, \tag{5.2}$$

having

$$\frac{\partial D(N, P)}{\partial P} = \ln \frac{1 - P/N}{P/N},$$

which is a monotonically decreasing function from $+\infty$ (at $P = 0$) to $-\infty$ (at $P = N$) vanishing at $P = N/2$. From the general property of logarithm function

$$\ln x \leqslant x - 1, \tag{5.3}$$

it follows that

$$-\sum_{i=1}^{N} (1 - f_i) \ln(1 - f_i) \leqslant P(f); \tag{5.4}$$

hence,

$$-P \ln P + P \ln N \leqslant D(N, P) \leqslant P + P \ln N - P \ln P. \qquad (5.5)$$

That is $D(N, P)$ is a monotonically (concave) increasing function of N vanishing at $N = P$ and logarithmically diverging for $N \to \infty$. For a very large N, or equivalently, for a sufficiently small P/N, $D(N, P)$ can be approximated as

$$D(N, P) \simeq P \ln N, \qquad (5.6)$$

and the finite variation $D(N + 1, P) - D(N, P)$ with the partial derivative

$$\frac{\partial D(N, P)}{\partial N} = \ln \frac{N}{N - P}.$$

From (5.4) we obtain the following *upper bound* for $d(f)$

$$d(f) \leqslant H(f) + P(f), \qquad (5.7)$$

where $H(f) \equiv -\sum_{i=1}^{N} f_i \ln f_i$, and the *lower bound*

$$d(f) \geqslant 2m(f), \qquad (5.8)$$

with $m(f)$ defined by (4.6).

The lower bound (5.8) is only the first order approximation of a series expansion of $d(f)$ in terms of the moments $M_h(f)$ $(h = 1, 2, \ldots)$. More precisely the following proposition holds.

PROPOSITION 5.1. *The logarithmic entropy $d(f)$ of a fuzzy set f can be expressed in terms of the moments by the following series expansion:*

$$d(f) = 2 \sum_{h=1}^{\infty} \frac{M_h(f)}{h}. \qquad (5.9)$$

The proof of the proposition is a direct consequence of the following series expansion of Shannon's function:

$$S(x) = \sum_{n=1}^{\infty} \frac{x^n(1 - x) + (1 - x)^n x}{n}, \qquad |x| \leqslant 1.$$

It is, moreover, easy to verify that the series

$$2 \sum_{n=1}^{\infty} \frac{z^n}{n} M_n(f), \qquad |z| \leqslant 1$$

is convergent to the function

$$G(z, f) = -\sum_{i=1}^{N} [f_i \ln[1 - z(1 - f_i)] + (1 - f_i) \ln(1 - zf_i)], \quad (5.10)$$

so that

$$M_n(f) = \frac{1}{2(n-1)!} \frac{d^n G}{dz^n} \bigg|_{z=0} \qquad (n = 1, 2, \ldots) \qquad (5.11)$$

and

$$d(f) = G(1, f). \qquad (5.12)$$

In such a way, by means of G and (5.12) and (5.11), we can generate the *entropy* and the *moments* of f. In order to have a unique function generating the entropy, the power, the energy and the moments it suffices to change (5.10) into

$$\hat{G}(w, z, f) = -\sum_{i=1}^{N} \{f_i \ln[1 - z(1 - f_i)] + (1 - f_i) \ln(1 - zf_i) - w_i f_i\}. \quad (5.13)$$

We get

$$\hat{G}(w, 0, f) = E(f), \qquad \hat{G}(1, 0, f) = P(f)$$

$$\hat{G}(w, 1, f) - \hat{G}(w, 0, f) = d(f),$$

$$\frac{1}{2(n-1)!} \frac{d^n \hat{G}(w, z, f)}{dz^n} \bigg|_{z=0} = M_n(f) \qquad (n = 1, 2, \ldots).$$

6. Maximum of the Uncertainty

In this section we wish to calculate the maximum of the uncertainty u by imposing suitable constraints on the energies, powers or effective powers of the fuzzy sets. We shall assume the L function in (3.4) coincident with the logarithm function. To realize the relevance of such a problem let us start maximizing the logarithmic entropy $d(f)$ of a fuzzy set keeping the energy and the power equal to two constants E and P, respectively. The maximum of $d(f)$ is reached for f equal to the fuzzy set

$$f_i = \frac{1}{1 + e^{\lambda w_i - \mu}} \qquad (i = 1 \cdots N), \tag{6.1}$$

λ and μ being constants whose values depend on E and P. This formula is identical to the *Fermi–Dirac distribution law* for a system of noninteracting particles when the spin is half-odd integral, in which case f_i is the "probability" that a state of energy w_i is occupied, μ/λ is the so-called *chemical potential* and λ is proportional to the inverse of the absolute temperature. In the case $w_1 = w_2 = \cdots = w_N = w$, then $E = wP$ and

$$f_i = \frac{P}{N} = \frac{E}{wN} \qquad (i = 1,..., N), \tag{6.2}$$

that is the maximum of $d(f)$ is reached when the energy E is *equidistributed* on the N elements, finding again the result expressed in Proposition 4.1.

More generally, let us consider the problem of maximizing the uncertainty $u(f^0,..., f^{M-1})$ under the constraints

$$\sum_{j=0}^{M-1} E(w^j, f^j) = E; \qquad \sum_{j=0}^{M-1} P(\phi^j, f^j) = P, \tag{6.3}$$

where $E(w^j, f^j)$ and $P(\phi^j, f^j)$ are the energy and the effective power of the fuzzy set f^j, respectively. The fuzzy sets that maximize u can be expressed, using the Lagrange's multipliers method, as

$$f_i^j = e^{\nu_i - 1 - w_i^j + \mu\phi^j} \qquad (i = 1 \cdots N; j = 0,..., M-1), \tag{6.4}$$

ν_i, λ, and μ being constants whose values depending on E and P, have to be determined by the equations

$$\nu_i = 1 - \ln Z_i \qquad (i = 1,..., N) \tag{6.5}$$

$$-\frac{\partial \ln Z}{\partial \lambda} = E, \qquad \frac{\partial \ln Z}{\partial \mu} = P, \tag{6.6}$$

where $Z_i = \sum_{j=0}^{M-1} e^{-\lambda w_i^j + \mu\phi^j}$ and $Z = \prod_{i=1}^{N} Z_i$. If λ, μ is a solution of (6.6), $u_{\max}$ can be expressed as

$$u_{\max} = \ln Z + \lambda E - \mu P. \tag{6.7}$$

Consider now the following particular cases:

(i) $\phi^j = 1$; $w_i^j = w^j$, that is P is equal to the total power and the weights of the energies depend only on the considered fuzzy set but not on the elements of I. From (6.4) it follows that

$$f_i^j = e^{\nu - w^j} \qquad (j = 0,..., M-1), \tag{6.8}$$

with $\nu = -\ln \sum_{j=0}^{M-1} e^{-\lambda w^j}$. Equation (6.8) coincides with the *Maxwell–Boltzmann* distribution for a gas of noninteracting molecules.

(ii) $\phi^j = j$, $w_i^j = jw_i$, that is the weight function of the energy for the jth fuzzy set is j times the one of f^1; moreover the weights of the effective powers on (6.3) are equal to the order numbers of the fuzzy sets.

In this case a meaningful quantity is $n_i = \sum_{j=0}^{M-1} jf_i^j$; indeed it allows to write (6.3) as

$$\sum_{i=1}^{N} n_i w_i = E; \qquad \sum_{i=1}^{N} n_i = P \tag{6.9}$$

so that $n_i w_i$ and n_i have the meaning of the *total energy* and *total effective power* of the ith element of I. We then obtain that in condition of maximum uncertainty, under the constraints (6.9), n_i is given by

$$n_i = \frac{-Me^{M\gamma_i} + e^{-\gamma_i} + (M-1)\, e^{-(M+1)\gamma_i}}{(1 - e^{-M\gamma_i})(1 - e^{-\gamma_i})} \qquad (i = 1,..., N), \tag{6.10}$$

where $\gamma_i \equiv \lambda w_i - \mu$. Equation (6.10) coincides with the "intermediate statistics" of Gentile (1940). As particular cases of (6.10), one obtains for $M = 2$ and $M = \infty$ the Fermi–Dirac and the Bose–Einstein distributions.

It is worthwhile to note the possibility of deducing, in an easy and unique manner, the previous classical and quantum distributions in a context of decision theory. In such a theory the above distributions have an unusual interpretation, in particular, as it is known (Münster, 1969), the intermediate statistics have no physical meaning if $2 < M < \infty$, since the eigenfunctions of quantum particles can only be symmetric or antisymmetric. If one interpretes f_i^j as *probability* that the ith element of I enjoys the property P^j, then the previous distributions have a statistical meaning. In the physical-statistic interpretation f_i^j is, in the derivation of the Maxwell–Boltzman distribution, the probability that the ith particle is in the jth level of energy w^j, and in the case of the intermediate statistics the probability that a level of individual energy w_i is occupied by a group of j particles ($j = 0, \ldots, M - 1$); n_i is the *average* number of particles at the level w_i. The maximum of the uncertainty corresponds to the *equilibrium situation* and the constraints (6.9) express the *conservation of the total energy and particle number.*

7. Some Applications to Decision Theory

In this section we shall refer to the particular case of two orthogonal properties only; however, the arguments we develop can be easily extended to the general case. As we have seen, the uncertainty in making decisions, in the case of two orthogonal measurable properties P and $\sim P$, reduces to the entropy $h(f)$ of the fuzzy set f obtained by considering the *measure function* of one of the two properties. We have also seen that a macroscopical description of a fuzzy set f can be obtained by quantities such as "entropy," "power," and "energy." We shall now show how it is possible to use these quantities in decision theory.

A decision may be regarded as any transformation of a fuzzy set into a classical set. Since after the decision has been taken the uncertainty vanishes, such a transformation yields a *quantity of information* that can be measured by $h(f)$ (De Luca and Termini, 1972).

A decision is generally determined by a set of rules by which one is able to transform $\mathscr{L}(I)$, the class of all fuzzy sets defined on I, or only some subsets of it, into classical sets.

If $D(I)$ is a subset of $\mathscr{L}(I)$ (in particular $\mathscr{L}(I)$ itself) and $C(I)$ the subset formed by the classical subsets of I, one may formally define a *decision* as any transformation F from $D(I)$ to $C(I)$:

$$F: D(I) \to C(I). \tag{7.1}$$

We distinguish between *deterministic* and *probabilistic* decisions; in the first case the decision rules uniquely determine the map F, whereas in the second case, they only determine the conditioned probability distributions that a decision transform a given fuzzy set into a classical set. Furthermore, decision rules may depend on "local" and/or "macroscopical" properties of the fuzzy sets, as for instance entropy, energy, power, etc. We shall see, although only in particular cases, that the previous quantities play an important role both in deterministic and probabilistic decisions.

I. *Deterministic Decisions*

A typical way of defining deterministic decisions is by a comparison of $f(x)$ with a given threshold θ $(0 \leqslant \theta \leqslant 1)$. Formally such a kind of decision can be defined by the transformation:

$$F(f, \theta)(x) = 1[f(x) - \theta], \qquad x \in I \tag{7.2}$$

where $1[x]$ is the unit step-function, $1[x] = 1$ if $x \geqslant 0$, $1[x] = 0$ if $x < 0$. More generally we can consider transformations in which θ is a function of x. In both cases the decision depends on local properties of f only. Furthermore, if for instance θ is itself a function of the entropy and/or energy of the given fuzzy set then the decision will depend on both *local* and *global* properties of f.

Going back to (7.2) let us introduce the following subsets of I:

$$F^1 = (x \in I \mid f(x) \geqslant \theta), \qquad F^0 = (x \in I \mid f(x) < \theta). \tag{7.3}$$

Let us now assume the number of elements of F^1 to be greater or equal to a fixed integer M $(0 \leqslant M \leqslant N)$. By (7.2) this implies that the fuzzy sets one has to consider are only those such that

$$P(f) \geqslant M\theta. \tag{7.4}$$

We now want to analyze whether there are some conditions that the entropy $h(f)$ has to satisfy in order to be $\#F^1 \geqslant M$.

In the case $\theta \geqslant 1/2$ one gets

$$\begin{aligned} h(f) &\leqslant \#F^1 T(\theta) + (N - \#F^1)\, T(1/2) \\ &\leqslant NT(1/2) - M[T(1/2) - T(\theta)] \leqslant NT(1/2). \end{aligned} \tag{7.5}$$

Therefore, in the case $\#F^1 \geqslant M$ the values of M and θ, $\theta \geqslant 1/2$, determine a *lower bound* for the energy and an *upper bound* for the entropy given by (7.4) and (7.5), respectively.

Let us now denote by $a(M, \theta)$ the function

$$a(M, \theta) = NT(1/2) - M[T(1/2) - T(\theta)]; \tag{7.6}$$

$a(M, \theta)$ is a linearly decreasing function of M assuming the minimum value at $M = N$ where $a(N, \theta) = NT(\theta)$. This quantity becomes as small as one wishes as θ approaches 1. Therefore, *if F^1 has to contain a large fraction of the elements of I then the fuzzy sets must have a suitable high power and be suitably sharp.*

In Table I are reported the values of the function $a(M, \theta)/N$ in the case $T(x) = S(x)$, for different values of the threshold θ and the fraction M/N. We get, for instance, that for $N = 10^3$ and $M = 900$ and $\theta = 0.95$

$$P(f) \geqslant 855; \qquad d(f) \leqslant 248.$$

In this case the upper bound 248 for the logarithmic entropy is less than half the absolute maximum which holds $\simeq 693$.

TABLE I

M/N θ	0.95	0.90	0.80	0.70	0.60	0.50
0.90	0.248	0.362	0.520	0.619	0.675	0.693
0.80	0.297	0.399	0.539	0.627	0.677	0.693
0.70	0.347	0.436	0.558	0.636	0.679	0.693
0.60	0.396	0.472	0.578	0.644	0.681	0.693
0.50	0.446	0.509	0.597	0.652	0.683	0.693
0.40	0.495	0.546	0.616	0.660	0.685	0.693
0.30	0.545	0.583	0.635	0.668	0.687	0.693
0.20	0.594	0.620	0.655	0.677	0.689	0.693
0.10	0.644	0.656	0.674	0.685	0.691	0.693

Let us now fix the number of elements of F^1 to be exactly equal to M. Then for the power, in addition to the lower bound (7.4), one also gets the upper bound:

$$P(f) \equiv \sum_{i \in F^1} f_i + \sum_{j \in F^0} f_j < (N - M)\theta + \sum_{i \in F^1} f_i$$
$$\leqslant (N - M)\theta + M. \tag{7.7}$$

Therefore, the power of a fuzzy set can vary only in the interval $[M\theta, (N - M)\theta + M]$.

For the entropy $h(f)$ we get the following limitations

$$h(f) \leqslant MT(\theta) + (N - M)\, T(1/2), \qquad \text{for} \quad \theta \geqslant 1/2$$

and

$$h(f) \leqslant (N - M)\, T(\theta) + MT(1/2), \qquad \text{for} \quad \theta \leqslant 1/2.$$

In the case of a constant power $P(f) = P$ we may always satisfy (7.4) by choosing a threshold θ such that

$$\theta \leqslant P/M.$$

If $\theta \geqslant 1/2$, which implies $P \geqslant M/2$, then $T(P/M) \leqslant T(\theta)$ and

$$a(M, \theta) \geqslant (N - M)\, T(1/2) + MT(P/M). \tag{7.8}$$

The right side of the previous formula gives a minimum for the upper bound of the entropy in the case of a constant power P and $\#F^1 > P$.

II. *Statistical Decisions*

a. We shall consider now *statistical decisions* performed by a universe of decision makers on a given fuzzy set f, with (stationary) probabilities which are related to the function f.

Let us introduce for any element of I the random variable f_i^* $(i = 1,..., N)$ that assumes the values 1 and 0 with probabilities p_i and $1 - p_i$. The most natural way of relating p_i to f_i is to assume

$$p_i = f_i \qquad (i = 1,..., N).$$

In such a case the (statistical) average of f_i is given by

$$\langle f_i^* \rangle = f_i\,, \qquad (i = 1,..., N)$$

and the *standard* deviation σ_i^2 by

$$\sigma_i^2 = f_i(1 - f_i), \qquad (i = 1,..., N).$$

The power $P(f^*)$ and the energy $E(f^*) = \sum_{i=1}^N w_i f_i^*$ of the classical set obtained after the decision are random variables whose *average values* are given by

$$\langle P(f^*)\rangle = P(f); \qquad \langle E(f^*)\rangle = E(f);$$

that is they coincide with the power and the energy of the fuzzy set f. If we suppose that the decisions performed on the elements of I are *statistically independent* then the variances $\sigma_P^2(f^*)$ and $\sigma_E^2(f^*)$ are given respectively by

$$\begin{aligned} \sigma_P^2(f^*) &= \sum_{i=1}^N f_i(1 - f_i) = m(f), \\ \sigma_E^2(f^*) &= \sum_{i=1}^N w_i^2 f_i(1 - f_i) \leqslant W^2 m(f), \end{aligned} \tag{7.9}$$

having denoted by W the maximum of $w_1\,,..., w_N$. So that recalling (5.8) the *variances of the energy* (*power*) *of* f^* *can be as small as one wants if* f *is sufficiently sharp*. Furthermore, the moments $M_n(f)$ $(N = 1, 2,...)$ appearing in the expansion (5.9) of the logarithmic entropy $d(f)$ coincide in this case with the sum of the absolute moments of order n of the random variable (Cramer, 1963)

$$\Psi_i \equiv f_i^* - \langle f_i^* \rangle = f_i^* - f_i .$$

Since $P(f^*)$ and $E(f^*)$ are the sum of N independent random variables, we consider now the case when the number N of the elements of I goes to the infinity in order to make use of asymptotic theorems of probability theory as the *central limit theorem.*

From the general theory (Cramer, 1963; see also De Luca and Ricciardi, 1967) it follows that $P(f^*)$ is normally distributed for $N \to \infty$ if and only if the series

$$\sum_{i=1}^{\infty} f_i(1 - f_i) \tag{7.10}$$

is divergent. Under this hypothesis the *probability density function* $\Phi(P)$ of $P(f^*)$ is given, for large N, by

$$\Phi(P) \simeq [2\pi m(f)]^{-1/2} \exp \frac{-(P - P(f))^2}{2m(f)} . \tag{7.11}$$

If we make the further hypothesis that, for all N, $0 < w \leqslant w_i \leqslant W$ ($i = 1,..., N$), $E(f^*)$ is also normally distributed having a probability density function $\Psi(E)$ given, for large N, by

$$\Psi(E) \simeq [2\pi\sigma_E(f^*)]^{-1/2} \exp - \frac{(E - E(f))^2}{2\sigma_E^2(f^*)} . \tag{7.12}$$

Since for all N,

$$d(f) \geqslant 2m(f) \geqslant (2/W^2)\, \sigma_E^2(f^*) \tag{7.13}$$

we have that *the energy and power of f determine the average values of the Gaussian* (7.11) *and the logarithmic entropy an upper bound for their variances*; in other words, the normal distributions of the energy and power of the classical set thus obtained are as "sharp" as one wishes if $d(f)$ is sufficiently small.

b. Let us now consider the case of only one decision maker D and a set I formed by a very large number N of elements; we suppose that the function f can assume only a finite number k of possible values in the interval [0, 1]: $\beta_1, \beta_2,..., \beta_k$. If $N_1, N_2,..., N_k$ are the numbers of elements assuming the values $\beta_1, \beta_2,..., \beta_k$ respectively, we have

$$P(f) = \sum_{i=1}^{k} N_i\beta_i , \qquad N = \sum_{i=1}^{k} N_i .$$

Under the hypothesis that $N_1, N_2, \ldots, N_k$ are very large, *let us furthermore suppose that D makes decisions on the elements of each group with probabilities* p_s *just equal to the values of* f; that is,

$$p_s = \beta_s \qquad (s = 1, \ldots, k).$$

p_s is, of course, the probability that D will decide 1 on the elements of the sth group. In the sth group the average statistical value of 1's is given by

$$n_s = p_s N_s = \beta_s N_s \qquad (s = 1, \ldots, k) \tag{7.14}$$

and the *standard deviation* by

$$\sigma_s^2 = N_s \beta_s (1 - \beta_s) \qquad (s = 1, \ldots, k). \tag{7.15}$$

Therefore, to the sth group of elements we associate a random variable x_s whose average value and standard deviation are given by (7.14) and (7.15). We can now consider the random variable X the sum of the x_s $(s = 1, \ldots, k)$:

$$X = \sum_{s=1}^{k} x_s ;$$

its average value and standard deviation are given by

$$\langle X \rangle = \sum_{s=1}^{k} N_s f_s = P(f),$$
$$\sigma^2 = \sum_{s=1}^{k} \sigma_s^2 = \sum_{i=1}^{k} N_i \beta_i (1 - \beta_i) = m(f). \tag{7.16}$$

(7.16) shows that the expected value of number of 1's produced coincides with the power of the fuzzy set and the statistical standard deviation with the first order moment $m(f)$.

8. Thermodynamical Analogies and Concluding Remarks

The aim of this section is to stress, even though in an intuitive more than formal manner, the analogies which exist between some concepts and the formalism developed in the previous sections, and the thermodynamics.

The concept of "entropy" plays a fundamental role in both thermodynamics and Shannon's *information theory*. The entropy of a finite scheme of events is identical to the physical one of a mechanical system, if the events are just the possible states of the system and the probability of an event is interpreted as probability of the system of occupying a state of suitable energy. However, in spite of this analogy a thermodynamical treatment of information processes in the context of Shannon's information theory has not been attempted, except recently in some particular cases (see, for instance, Beneš, 1963; Mandelbrot, 1970),[3] since a thermodynamic-like energy concept has not been defined.

The exigence of a thermodynamics for information theory was very deep in von Neumann (1966):

> "I have been trying to justify the suspicion that a theory of information is needed and that very little of what is needed exists yet. Such small traces of it which do exist, and such information as one has about adjacent fields indicate that, it found, it is likely to be similar to two of our existing theories: formal logics and thermodynamics. It is not surprising that this new theory of information should be like formal logics, but it is surprising that it is likely to have a lot in common with thermodynamics. . . ."

In the previous sections we have seen how it is possible to introduce in the setting of fuzzy sets theory some thermodynamic-like quantities such as *entropy*, *energy*, *power* and the role that they play in decision theory. Furthermore we have seen in Section 6 that the power and energy of a fuzzy set correspond to the number of particles and to energy of a physical system; the *condition of maximum uncertainty* corresponds to the *thermodynamical equilibrium*. In the end of the section we shall briefly outline some links existing between the maximum uncertainty conditions and the "learning processes;" we now want to show that by means of the concepts of entropy, energy, power. It is possible to develop a formalism similar to the one of thermodynamics, even though the interpretation of quantities is, generally, completely different from the physical one.

Let us denote by $U(E, P, x_\nu)$ the maximum uncertainty in making decisions keeping the total energy and effective power equal to the constant values E and P; $\{x_\nu\}$ denote the set of external parameters describing the system (as w_i^j or ϕ^j, N etc.). If the conditions are slightly changed but in a way to reach another condition of maximum uncertainty, we have

$$dU = \frac{\partial U}{\partial E} dE + \frac{\partial U}{\partial P} dP + \sum_\nu \frac{\partial U}{\partial x_\nu} dx_\nu \tag{8.1}$$

[3] V. E. Beneš (1963) introduced a thermodynamic theory of random traffic in connecting networks. In this theory the number of "calls in progress" in a state of the net is analogous to the energy of a physical system. B. Mandelbrot (1970) makes use of a thermodynamical approach to linguistics assuming that the "energy of a word" coincides with the "number of letters" in the word.

or

$$dE = \tau\, dU + \mu\, dP - \sum_{\nu} X_{\nu}\, dx_{\nu}, \tag{8.2}$$

having set

$$\frac{\partial U}{\partial E} = \frac{1}{\tau}, \qquad \frac{\partial U}{\partial P} = \frac{-\mu}{\tau}, \qquad \frac{\partial U}{\partial x_{\nu}} = \frac{X_{\nu}}{\tau}. \tag{8.3}$$

Equation (8.2) formally coincides with the *first law of thermodynamics* on the condition that τ is interpreted as *absolute temperature*, μ as *chemical potential*, $\tau\, dU$ as *heat* and the terms $-\mu\, dP$ and $\sum_{\nu} X_{\nu}\, dx_{\nu}$ as the *chemical* and *external elementary works*, the latter caused by a change in the external parameters, which yield changes of the total (internal) energy. If the only external parameter that can be changed is $x_{\nu} = N$, and supposing the change of N to be very small with respect to N, then N is the correspondent of the *volume* and $X_{\nu} = \partial U/\partial N$ of the *pressure*.

It is very important, before any attempt at the interpretation of the quantities formally introduced by (8.3), to give an interpretation of the "conditions of maximum uncertainty" which, as said before, physically correspond to thermodynamical equilibrium.

We shall refer in the following, for simplicity, to the case of a single fuzzy set and consider the *normalized entropy* (4.2).

If we make the next decisions on a fixed fuzzy set define in I, if the entropy is not increasing with the number of decisions, the *normalized entropy is on the contrary nondecreasing*. Let us consider now a procedure in which one supposes that a decision maker on a set I of objects described by a fuzzy set, changes, after any decision, the membership function according to some rules which depend on whether the decision made is considered "correct" or "wrong." Let us stress that "correct" and "wrong" have a meaning only with respect to a given "teacher system" that is a given way of deciding the objects of I.

Let us consider as teacher a system T whose decision function is a threshold function $1[f(x) - \theta]$, $0 \leqslant \theta \leqslant 1$ and a decision maker L. L behaves according to the following decision rule. Let us suppose that L at the mth step decides in correspondence of a certain value $f(x_0)$ of the membership function; if the decision is 1 and is right [wrong] then L will decide in the same way, e.g. 1 [in the opposite way, e.g. 0] at the step $m + 1$ for all the other objects of I, such that $f(x) \geqslant f(x_0)[f(x) \leqslant f(x_0)]$.

If the decision of L is 0 and right [wrong] then L will decide at the $(m + 1)$th step in the opposite way [in the same way] for all the other objects for which

$f(x) \leqslant f(x_0)[f(x) \geqslant f(x_0)]$. In all the other cases the membership function will remain unchanged.

Using the previous decision rule, after a suitable number of steps, it will be up to L to decide only a subset I^* of I formed by objects for which $f^* \simeq \theta$.

The power of f^* is then given by

$$P(f^*) = \sum_{i \in I^*} f_i^* \simeq k\theta \equiv P,$$

where $k = \# I^*$. The entropies $h(f^*)$ and $\nu(f^*)$ will be given by

$$h(f^*) \simeq kT(\theta) = kT(P/k); \qquad \nu(f^*) = T(P/k),$$

that is after a suitable number of steps the entropy and the normalized entropy will reach the maximum value with respect to all the fuzzy sets with a power equal to $P = k\theta$, which are the only allowed by the decision rule under consideration.

At an intuitive level the "learning processes" in the context of fuzzy sets theory, have to be such as to make the "best" possible use of the "information" deriving from the responses of the teacher. In this way a "learning system," as in the example we gave, has to be able, from a given response, to decide all the less ambiguous situations. This fact produces a very rapid decreasing in the total uncertainty, but after a suitable number of steps the average uncertainty tends to a constant value, and this because the situations which remain to be decided are the most ambiguous ones among those which are allowed by the interaction *learning system–teacher*. A deeper mathematical and conceptual analysis of these processes is needed and can be the subject of future papers. We limit ourselves here to stressing, on the one hand, the noteworthiness of the possibility of characterizing the learning processes by means of a *variational principle* (on the average uncertainty) similar to the ones of classical mechanics, and, on the other hand, the possibility of using a formalism as the one of thermodynamics which will certainly help in a specification of typical features of learning processes.

Acknowledgments

The authors are indebted to Dr. G. Svetlichny, Dr. A. Restivo, and Dr. S. Termini for their useful comments and to Professor E. R. Caianiello and Professor M. P. Schützenberger for their encouragement and suggestions.

Received: March 21, 1973

References

Beneš, V. E. (1963), A "thermodynamic" theory of traffic in connecting networks, *Bell System Tech. J.* **47**, 567–607.

Capocelli, R. M., and De Luca, A. (1972), Fuzzy sets and decision theory, *Notices Amer. Math. Soc.* **19**, A.709, abstract.

Capocelli, R. M., and De Luca, A. (1972), Measures of uncertainty in the context of fuzzy sets theory, Atti del II° Congresso Nazionale di Cibernetica di Casciana Terme, Pisa.

Cramer, H. (1963), "Random variables and probability distributions," Cambridge University Press, Cambridge.

De Luca, A., and Ricciardi, L. M. (1967), Formalized neuron: Probabilistic description and asymptotic theorem, *J. Theoret. Biol.* **14**, 206–217.

De Luca, A., and Termini, S. (1970), Algebraic properties of fuzzy sets, *Notices Amer. Math. Soc.* **17**, 944, abstract; (1972) *J. Math. Anal. Appl.* **40**, 373–386.

De Luca, A., and Termini, S. (1971), Algorithmic aspects in complex systems analysis, *Scientia* **106**, 659–671.

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.

Hardy, G. .H, Littlewood, J. E., and Pólya, G. (1967), "Inequalities," Cambridge University Press, Cambridge.

Khinchin, A. I. (1957), "Mathematical Foundations of Information Theory," Dover Publ., New York.

Gentile, G. (1940), Osservazioni sopra le statistiche intermedie, *Nuovo Cimento* **17**, 493–497.

Mandelbrot, B. (1970), The concept of temperature of discourse, *in* "A Critical Review of Thermodynamics" (E. B. Stuart, B. Gal-or, and A. J. Brainard, Eds.), Mono Book Corp., Baltimore, MD.

Münster, A. (1969), "Statistical Thermodynamics," Springer-Verlag and Academic Press, New York.

Neumann, v. J. (1966), "Theory of Self-Reproducing Automata," University of Illinois Press, Urbana, IL/London.

Shannon, C. E., and Weaver, W. (1962), "The Mathematical Theory of Communication," University of Illinois Press, Urbana, IL.

Vajda, I. (1969), A contribution to the informational analysis of pattern, *in* "Methodologies of Pattern Recognition," pp. 509–519, Academic Press, New York.

Wald, A. (1971), "Statistical Decision Functions," Chelsea Publ. Company, New York.

Zadeh, L. (1965), Fuzzy sets, *Information and Control* **8**, 338–353.