STOCHASTICA Vol. VI nº 3 (1982)

FUZZY SETS AS SET CLASSES

Ton Sales

In remembrance, Joseph Kampé de Fériet (1893-1982).

0. Introduction.

Fuzzy sets have been studied in various forms. We now offer a presentation of fuzzy sets whereby they are conceived as representatives of a whole class of sets (that are themselves subsets of the universe of objects on which the fuzzy set is defined). Such an approach offers certain obvious advantages for a generalized theory of truth, in whose setting the approach arose. These advantages are not explored here. We mention here, however, the intuitive appeal such a theory may exert from the philosophical side. It allows, in fact, to view vague concepts and imprecise observations -such as the ones fuzzy set models try to formalize- as the compaction of a series of classical standard sets representing sharp-edged concepts or observations (as e.g. in ostension).

The approach expounded here derives from a development in Dempster [v.5] which interestingly was formalized and systematized in Shafer [5]. It was Kampé de Fériet's merit to treat Shafer's "probabilities" as measured by a set of observers who were themselves valued additively by some super-observer. He was also the first to point out possible links between this set of measures and Zadeh's membership grade for fuzzy sets ([8]).

The following lines try to extend Kampé de Fériet's scheme to cover the present author's ideas about fuzzy sets as representatives of set classes.

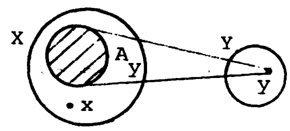
1. Set Classes.

Suppose a finite set X ("universe") made up of "objects" x. Subsets $A \subset X$ will be called here "sets" (or "properties"). Any collection of sets $A \in P(X)$ forms a class $\mathcal{A}$ that is not necessarily a set (since it may contain repeated occurrences of a given set). $\mathcal{A}$ will be sometimes called set sample. Throughout we suppose $\mathcal{A}$ is finite. Each set in the sample will be noted A_y (where y is the index that identifies each set occurrence).

As is known, a set class can be pictured as a set of pairs (y, A_y). If we note by Y the (finite) set of indices (we call positions), this set of pairs define the mapping $\pi: Y \to P(X)$, we call perspective (on X), where the elements (y, A_y) -we call eventualities- are chosen so that they give exactly all the sets A_y in the sample $\mathcal{A}$.

Rangeπ contains all sets considered or observed in the sample. Range$\pi = P(X)$ means the whole universe X has been considered or observed. Alternately, $\pi(y)=A_y$ means the "perspective from y" or (in Kampé de Fériet's terms) "the set where a given $x \in X$ can be located (from y)". In the latter case the characteristic function of A_y performs the role of a "location rule": x is located (resp. not located) in $\pi(y)$ iff is (resp. is not) in A_y.

The above formal interpretation scheme is similar to the one proposed by Kampé de Fériet in 1980 ([3]). It allows for several useful characterizations of sets A of a universe X of objects, and also several measures on P(X), by means of a (separately chosen) universe Y of positions. Sets A are then perspectives from the positions of Y whereas the universe X has its elements locatable or not depending on the position from which they are considered of observed. In this approach, eventualities are just possi



ble pairings of sets A and positions y. As shown later, a given sample A induces in P(X) a probability measure, which in turn can be summarized in fuzzy set form.

To complete this section, we mention that the mapping $\sigma : A \rightarrow P(Y)$ such that $\sigma = \pi^{-1}|_A$ is what we call <u>selection</u> function. The set $\sigma(A)$ contains those positions from which the set A is considered or observed, or from which all elements $x \in A$ can be located.

2. Non-Additive Measures on Set Classes: Probabilities and chance.

Since P(Y) is a Boolean algebra structure (with operations $\cap, \cup$ and C), a valuation $v : P(Y) \rightarrow [0,1]$ can be defined on P(Y) such that

i) If $A \subset B$ then $v(A) \leqslant v(B)$ (Monotonicity)

ii) $v(\emptyset)=0$ and $v(Y)=1$

iii) for any $i \neq j$ if $A_i \cap A_j = \emptyset$ then $v(\cup A_i) = \Sigma v(A_i)$ (Additivity).

Thus v is technically an additive probability. It can be called "hierarchy of positions" and used to value any sample A. In the observational interpretation of Kampé de Fériet [3], the v function (he notes by $\underline{\lambda}$) represents the way a distinguished observer ("headquarters" o HQ) values positions; this valuation allows HQ to locate selectively an object x in given sets $\in P(X)$.

An additive probability $p: P(X) \to [0,1]$ is not aiways definable on X but, we now hypothesize, we can always define two subadditive probability measures on X. Following Shafer's usage ([3]), we shall hereafter sometimes call epistemic probability (or probability, for short) every sub-additive probability, and chance every additive probability.

We define now the outer measure $\sigma^*(A)$ and the inner measure $\sigma_*(A)$ of a set A of P(X). We suppose, as above, X and Y are given

$$\sigma^*(A) = \{y \mid A_v \cap A \neq \emptyset\}$$

$$\sigma_*(A) = \{y \mid \emptyset \neq A_y \subset A\}$$

While $\sigma^*(A)$ is the set of positions from which A is (at least partly) considered or observed, $\sigma_*(A)$ is the set of positions from which some part (or the whole) of A is considered or observed. In Kampé de Fériet's epistemic interpretation, $\sigma^*(A)$ and $\sigma_*(A)$ give the set of positions ("observers" in Kampé's terminology) from which the statement "$x \in A$" is, respectively, plausible and certain.

The following results are immediate:

$$\sigma^*(X) = \sigma_*(X) \quad \text{(hence we write } Y_x = \sigma^*(X) = \{y \mid A_y\})$$

$$\emptyset \subset \sigma_*(A) \subset \sigma^*(A) \subset Y_x.$$

If we now value (by the v function) the $\sigma^*(A)$ and $\sigma_*(A)$ positions we have:

$$p_*(A) = \frac{v(\sigma_*(A))}{v(Y_x)}$$

$$p^*(A) = \frac{v(\sigma^*(A))}{v(Y_x)} \quad (v(Y_x) \text{ is always non-zero}).$$

These two new measures is what we respectively call lower and upper (epistemic) probability or, for reasons that will later be apparent, belief and plausibility. Interesting properties are:

$$0 \leqslant p_*(A) \leqslant p^*(A) \leqslant 1$$

$$p_*(A) = 1-p^*(\bar{A}) \text{ and } p^*(A)=1-p_*(\bar{A}) \text{ (where } \bar{A} \text{ is } X-A)$$

or also:

$$p_*(A)+p^*(\bar{A}) = p^*(A) + p_*(\bar{A})$$

$$p_*(A)+p_*(\bar{A}) \leqslant 1 \leqslant p^*(A) + p^*(\bar{A}).$$

Now, the analogy of both measures to Shafer's belief (or "support") and plausibility is striking. Recall that Shafer defines belief in 1976 ([5]) as any function Bel: $P(X) \to [0,1]$ satisfying the following conditions:

i) $\text{Bel }(\emptyset) = 0$

ii) $\text{Bel }(X) = 1$

iii) for any $n \geqslant 1$, $\text{Bel }(\overset{n}{\cup} A_i) \geqslant \sum_i \text{Bel }(A_i) - \sum_{i<j} \text{Bel}(A_i \cap A_j) + \ldots$

$$+(-1)^{n+1} \text{Bel}(\overset{n}{\cap} A_i).$$

From (iii) follows

$$\text{Bel }(A) + \text{Bel }(\bar{A}) \leqslant 1$$

Moreover, by definition, we have

$$\text{Pl}(A) = 1-\text{Bel}(\bar{A})$$

Shafer's plausibility $\text{Pl}:(X) \to [0,1]$ satisfies instead

i) $Pl(\emptyset) = 0$

ii) $Pl(X) = 1$

iii) for any $n \geq 1$, $Pl(\overset{n}{\cup}A_i) \leq \sum_i Pl(A_i - \sum_{i<j} Pl(A_i A_j) + \ldots$
$+ (-1)^{n+1} Pl(\overset{n}{\cap}A_i)$.

From (iii) follows

$$Pl(A) + Pl(\bar{A}) \geq 1$$

Indeed, p_* and p^* satisfy exactly the same conditions as Shafer's Bel and P functions. So we feel justified in identifying p_* with Bel and p^* with Pl and giving hereafter p_* and p^* the names belief and plausibility. Further support for it is the fact that $p_*(A)$ and $p^*(A)$ is, as we told before, the proportion of positions from which the statement "$x \in A$" is deemed <u>certain</u> and <u>plausible</u>, respectively.

Further properties are:

- Both p_* and p^* are <u>fuzzy measures</u> in Sugeno's sense ([6]).
- Zadeh's possibility measure ([8]) is, if normalized (ie. $\sup_{x \in X} f(x)=1$), indeed a particular case of the plausibility p^*.
- p_* and p^* coincide for, and only for, those sets which are a partition of X and which are in one-to-one correspondence with the positions of Y. Then and only then, $p_*(=p^*)$ is an <u>additive probability</u> or <u>chance.</u>

3. From Set Class to Fuzzy Set.

We now explain how the probability distribution induced on X by a set class or sample of sets A_y defines univocally a fuzzy set $\underset{\sim}{A}$, which is the summary and representative -though not the only possible one- of class *A*.

Applying above formulas to the particular case where the sets $A \subset X$ are singletons we have

$$A = \bigcup_{x \in A} \{x\} \quad \text{and} \quad X = \bigcup_{x \in X} \{x\}$$

$$\sigma^*(\{x\}) = \{y \mid x \in \pi(y)\}.$$

From both we have:

$$\sigma^*(A) = \bigcup_{x \in A} \sigma^*(\{x\}) \text{ and } Y_x = \sigma^*(x) = \bigcup_{x \in X} \sigma^*(\{x\})$$

Let us recall that a fuzzy set $\underset{\sim}{A}$ is any set of objects x from a reference universe X for which there is a function (called "characteristic function") $A = X \to [0,1]$ assigning every $x \in X$ a value $A(x) \in [0,1]$ named membership grade (of x in $\underset{\sim}{A}$), in a generalization of the classical notions of set and characteristic function. (For convenience we use "A(x)" instead of the more appropriate "$\mu_{\underset{\sim}{A}}(x)$").

We now define the membership grade A(x) of an object $x \in X$ in a fuzzy set $\underset{\sim}{A}$ as the plausibility (=upper probability) p^* of the fact that x belongs to it, i.e. the plausibility associated to {x} induced by the sample *A* of sets A_y defined in X. In symbols:

$$\underline{A(x) = p^*(\{x\}),} \text{ that is: } A(x) = \frac{v(\sigma^*(\{x\}))}{v(Y_x)}$$

Clearly A(x) inherits from p^* the property of depending on the point (object) x and of taking values in the [0,1] interval. It is therefore a "characteristic function" $A: X \to [0,1]$ of the required type, and so it defines automatically a fuzzy set $\underset{\sim}{A}$ (in X). The value of A(x) in each point, now conceived as membership grade takes on the additional epistemic connotation of "plausibility" or "upper probability" induced by a set sample *A*.

If we note by $A_y = X \to \{0,1\}$ (where $A_y(x) = 1$ if $x \in A_y$ and zero otherwise) the (classical) characteristic function of each set A_y in the sample, we have as a result:

$$A(x) = \frac{\Sigma_Y\ v(y) . A_y(x)}{\Sigma_Y\ v(y)}$$

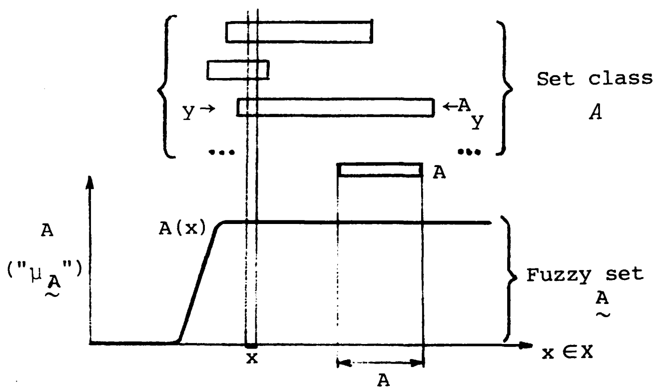
This formula characterizes the membership grade A(x) of each element in a fuzzy set as the (v-) weighted mean of all values of the characteristic function $A_y(x)$ that all sets in the sample take in this point x.

In the particular case where all positions are equal-valued, we have:

$$A(x) = \frac{\Sigma_Y\ A_y(x)}{\text{card } Y}$$ that we can set $A(x) = \frac{a}{n}$, for short.

Thus A(x) can be viewed as a mean of cases, or a relative cardinal.

The following illustration shows how a fuzzy set $\underset{\sim}{A}$ is built up on a sample *A* of sets A_y.



Interestingly, this characterization is essentially compatible with those given by Black in 1937 ([1]), De Luca-Termini in 1977 ([2]) and Kampé de Fériet in 1980 ([3]). What for Black is the proportion of persons believing "A_x" (or positive answer to the question "A_x?") with respect to the totality (Black writes "$N \to \infty$") of surveyed persons (or of responses), for the others is the proportion of "positive decisions" (De Luca-Termini) or of "positive (experimental) tests" (Kampé de Fériet) -for a given x- with respect to the totality of the $|Y|$ possible decisions or of the $|Y|$ tests. Leaving aside De Luca-Termini's rather linguistic and formal bias and Kampé de Fériet's clearly epistemological setting, the above formulas obviously characterize the membership grade as a proportion of occurrences of an object x in the sets A_y of a sample $\mathcal{A}$.

This notion of "occurrences of an object" in the frame of possible occurrences suggests an analogy with the contingency of a modal proposition in the frame of possible worlds semantics by Kripke et al. Occurrences or eventualities clearly correspond to possible worlds, but with a difference: they are always defined with respect to a reference sample.

We add that, after De Luca-Termini's 1977 definition ([2]), the following function measures the "unexpectedness" of the A_y "decisions":

$$I(A_y) \overset{def}{=} \sum_{x \in X} [A_y(x) . L(\frac{1}{A(x)}) + (1-A_y(x)) . L(\frac{1}{1-A(x)})]$$

where L is a continuous concave function in the $[1,+\infty)$ interval (e.g. logarithm). We quote this definition because authors say it allows a direct calculation of the entropy $u(\underset{\sim}{A})$ of a fuzzy set $\underset{\sim}{A}$ via the arithmetic mean

$$u(\underset{\sim}{A}) = \underset{y \in Y}{\text{mean}} \text{ of the } I(A_y)\text{'s}$$

For all we said, a fuzzy set can be conceived as the information summary -or the representative- of a (v-valued) set class or sample $\mathcal{A}$. This relationship is not one-to-one: a given fuzzy set is usually the representative for a whole family of set classes $A_1,\ldots,A_i,\ldots$

$$\begin{array}{c} A_1 \ldots A_i \ldots \\ \searrow \quad \swarrow \\ \underset{\sim}{A} \end{array}$$

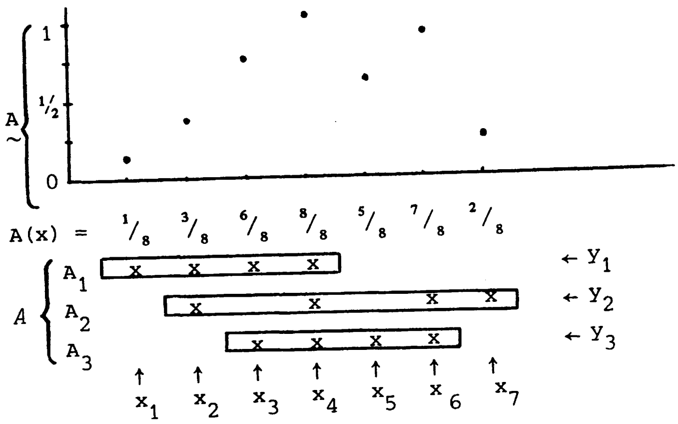
4. From Fuzzy Set to Set Class.

We now approach the problem from the opposite side. An easily provable result is this: Given a fuzzy set $\underset{\sim}{A}$ (in a finite universe X) we can find: (a) a finite set Y of positions, and (b) an additive valuation v on Y, both Y and v generally not unique, such that, for any $y \epsilon Y$, p $(\{x\}) = A(x)$; moreover, once Y and v found we can also find (c) a class A (in general not unique) of (possible repeated) sets A_y such that, for any $x \epsilon X$,

$$A(x) = \frac{v(\{y \mid x \epsilon A_y\})}{v(Y)} .$$

As an example, suppose the membership function A(x) of $\underset{\sim}{A}$ is given pointwise by the following figure and has eight different values. A ternary position set $Y = \{y_1, y_2, y_3\}$ suffices, as well as the valuation instance listed below, to define a ternary sample $A=\{A_1,A_2,A_3\}$ in the way depicted in the figure.

$\nu(\emptyset)$	$= 0$		$A(x) = 0$	corresponds	to	$\emptyset$
$\nu(\{y_1\})$	$= 1/8$		$A(x) = 1/2$	"	"	$\{y_1\}$
$\nu(\{y_2\})$	$= 2/8$		$A(x) = 2/8$	"	"	$\{y_2\}$
$\nu(\{y_3\})$	$= 5/8$		$A(x) = 3/8$	"	"	$\{y_1 y_2\}$
$\nu(\{y_1 y_2\})$	$= 3/8$	$\rightarrow$	$A(x) = 4/8$	"	"	-
$\nu(\{y_1 y_3\})$	$= 6/8$		$A(x) = 5/8$	"	"	$\{y_3\}$
$\nu(\{y_2 y_3\})$	$= 7/8$		$A(x) = 6/8$	"	"	$\{y_1 y_3\}$
$\nu(\{y_1 y_2 y_3\})$	$= 1$		$A(x) = 7/8$	"	"	$\{y_2 y_3\}$
			$A(x) = 1$	"	"	y



Now we propound a simple method, that can be used in certain conditions, to obtain the class A directly from $\underset{\sim}{A}$, using the decomposition theorem. The latter, also called "resolution theorem" by Zadeh, states that every (finite) fuzzy set $\underset{\sim}{A}$ can be expressed in this way:

$$\underset{\sim}{A} = \overset{M}{\cup} \alpha_i \, A_{\alpha_i}$$

where M is the number of different values of A(x) effectively found in $\underset{\sim}{A}$.

The proposed method consists of just choosing a Y with $|Y| = M$ positions; if the M values are not equal-distanced, n equal-distanced values are chosen. Suppose we fix n eventualities. We then choose tha A_y's so that they coincide with the M sets A_{α_i}. We shall have

$$A_y = A_{\alpha_i}$$

$$w_y = w_i \quad \text{where } w_i = n(\alpha_i - \alpha_{i-1})$$

(w_y is the number of repeated occurrences of each A_y, w_i is the number of occurrences of A_{α_i}).

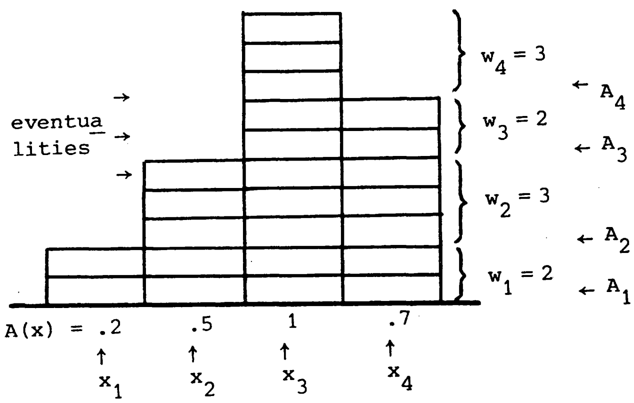
Consider an easy example. Suppose $\underset{\sim}{A}=[\overset{x_1}{0.2}\ \ \overset{x_2}{0.5}\ \ \overset{x_3}{1}\ \ \overset{x_4}{0.7}]$. By virtue of the decomposition theorem we can write:

$$\underset{\sim}{A} = 0\cdot 2\cdot\underset{A_1}{[1\ 1\ 1\ 1]} \cup 0\cdot 5\cdot\underset{A_2}{[0\ 1\ 1\ 1]} \cup 0\cdot 7\cdot\underset{A_3}{[0\ 0\ 1\ 1]}\quad 1\cdot\underset{A_4}{[0\ 0\ 1\ 0]}$$

Let us choose an $n(=|Y|)$ such that all products $\alpha_i n$ give integer values; we obtain $n=10$ (and thus $|Y|= 10$ eventualities). We also obtain $w_1 = 2$, $w_2 = 3$, $w_3 = 2$ and $w_4 = 3$. The resulting class is, therefore:

$$A = \{2A_1\text{'s},\ 3A_2\text{'s},\ 2A_3\text{'s},\ 3A_4\text{'s}\}$$

Graphically,



5. The Coherency Condition.

Unfortunately, a <u>fuzzy set</u> does not correspond to a single <u>set class</u> but to many, so both formalisms are neither equivalent now convey the same amount of information. In one case, though, a standard form can be given for fuzzy sets so that the set classes a fuzzy set represents are fewer, or even reduce to one single class. The condition that gives this result is called, follwing De Luca-Termini [2], "coherency condition". It may be stated in this way:

(Coherency Condition:) If, for any y and y' of Y and any x of X, we always have $A_y(x) \leqslant A_{y'}(x)$ or $A_y(x) \geqslant A_{y'}(x)$, then the class $\mathcal{A}$ of the A_y sets is a coherent class.

The coherency condition ("CC", for short, from now on) is nothing but the formalization of an implicit hypothesis, that of assuming that each position y is the site of an observer with a "coherent" behavior. This supposition is often unrealistic in practice, since the sets in the sample do not actually present themselves in the form the CC requires. But wherever it can be safely assumed, the CC gives the fuzzy set some adequate properties. In a notation more closely related to De Luca-Termini, the CC can be expressed as follows:

"For any y and y' in Y, either $\delta_y(x,A(x)) \leqslant \delta_{y'}(x,A(x))$ or vice versa".

The CC amounts to say that (a) there exists a total ordering of the sets A with respect to set inclusion, and (b) there exists a total ordering of the positions y in Y. Both orderings are, moreover, well-orderings.

If the CC is satisfied, we have the following interesting properties:

1) The plausibility p^* on X is a possibility measure (in Zadeh's sense); so then we have, for instance:

$$p^*(A \cup B) = \max\ (p^*(A),\ p^*(B))$$

$$p^*(A \cap B) \leqslant \min\ (p^*(A),\ p^*(B))$$

2) The belief p_* on X (defined by $p_*(A) = 1 - p^*(\bar{A})$) has the properties:

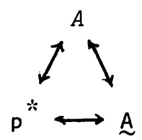
$$p_*(A) = \inf_{X-A} \bar{A}(x), \text{ and also:}$$

$$p_*(A \cap B) \geqslant \max\ (p_*(A),\ p_*(B))$$

$$p_*(A \cap B) = \min\ (p_*(A),\ p_*(B)).$$

3) if v assigns all positions y equal values then the CC can be expressed in the slightly restrictive way of De Luca-Termini ([2]); moreover, the class A is then <u>unique.</u>

If the CC and the above (3) condition hold, we have three substantially equivalent formalisms which correspond one-to-one among themselves: a set class or sample A (of subsets A_y of x), a possibility distribution p^* (on X) and a fuzzy set $\underset{\sim}{A}$ (with $A: X \to [0,1]$) defined on X. Graphically,



6. Some final considerations.

Often a concept (summarizingly represented by $\underset{\sim}{A}$) derives from an ostension process, in which a sample A of several (possibly repeated) sets A_y allowing yes/no binary membership decisions are effectively presented to us. We then submit them to a valuation v in accordance with force, frequency or representativeness of the set, or context. The result is the compaction of all perceived information into some abridged $\underset{\sim}{A}$-type form. This $A \to \underset{\sim}{A}$ process is typical of <u>abstraction</u>, so much of inductive abstraction from observational yes/no results as of formal or symbolic abstraction where several (often merely hypothesized) sharp-edged A_y's are manipulated and combined to get some smoother-edged, more general concept.

Sometimes the abstraction process is reversed and gives way to the inverse $\underset{\sim}{A} \to A$ process which we could call "concretion" (as opposed to "abstraction") or, after Artificial Intelligence terminology, "instantiation". (where the A_y's in A play the role of elementary "frames") (see Minsky [4]). As we found, this decomposition process is not unique, save when the "concept" $\underset{\sim}{A}$ is structured in the "coherent" way we described above and every "frame" A_y is assigned equal value, a condition that cannot always be met in real-world conditions.

References.

[1] BLACK, M., "Vagueness", Phil. Sci. 4 (1937), 427-455.

[2] De LUCA, A. and S. TERMINI, "Measures of Ambiguity in the Analysis of Complex Systems", in J. Gruska (ed.) Mathematical. Foundations of Computer Science, Springer 1977.

[3] KAMPÉ DE FÉRIET, J., "Une interprétation des mesures de plausibilité et de crédibilité au sens de G. Shafer et de la fonction d'appartenance définissant un ensemble flou de L. Zadeh" Pub. IRMA Vol. 2, fasc. 6, n°. 2, Lille 1980.

[4] MINSKY, M., "A Framework for Representing Knowledge", in P.H. Winston (ed.). The Psychology of Computer Vision, McGraw-Hill 1974.

[5] SHAFER, G., A Mathematical Theory of Evidence , Princeton University Press, 1976.

[6] SUGENO, M., Theory of Fuzzy Integrals and its Applications , Ph. D. Thesis, 1974.

[7] ZADEH, L. A., "Fuzzy Sets", Info. and Control 8 (1965), 338-353.

[8] ZADEH, L.A., "Fuzzy Sets as a Basis for a Theory of Possibility", Fuzzy Sets and Systems 1 (1977), 3-28

Remark.

Since the contents of this paper are a logic continuation of Kampé de Fériet's Information Theory reflections in the setting of Set Theory (and Fuzzy Set Theory), they bear a close relationship to parallel efforts such as the growing research in "random sets", as exemplified by the two representative papers cited below:

NGUYEN, H. T.: "On Random Sets and Belief Functions", UC Berkeley ERL Memo M77/14, 1977.

GOODMAN, I. R.: "Fuzzy Sets as Equivalence Classes of Random Sets" in R.R. Yager, ed.: Fuzzy Sets and Possibility Theory: Recent Developments, Pergamon 1982.

Facultat d'Informàtica.
Dulcet, 12
BARCELONA.