We study classical or generalized partitions of a given finite set from two points of view. On the one hand, we consider fuzzy binary relations $R$ defined on a finite set and $\ast$-transitive for a binary operation $\ast$, such that $d=1-R$ is a distance. We characterize such $\ast$-relations by matricial properties and we give several examples such as $m$-supermetricity which is a necessary condition for $m$-hypermetricity. Partitions are then constructed by means of cliques and spheres associated with the distance $d$. The case of fuzzy partitions is specially investigated and an example from coding theory is given, quasi-partitions are connected with $k$-ultrametricity. On the other hand, we characterize fuzzy partitions $\Delta$ with the help of an information associated with a given threshold $\eta$ which is the lower value of the membership functions defining the fuzzy classes of $\Delta$. By evaluating the probabilities of these classes, we construct a quantity which increases when the partition $\Delta$ is refined. The gain of information, resulting from the replacement of a given fuzzy partition by another one, decreases when we sharpen a fuzzy partition and it measures the "spread" of $\Delta$. It is then a good tool to choose a fuzzy partition in a family or to improve a given classification of the studied population, for example.


INTRODUCTION

Fuzzy binary relations between elements of a finite population $I$, generalizing classical deterministic relations, admit various applications in clustering or data analysis. In particular cases, they yield distances between the members of $I$. We consider such relations, which further verify a generalization of the transitivity property. We study extensions of hypermetricity and ultrametricity of the distances, looking for results concerning the existence of spheres and cliques in $I$, we compare several such distances, weaker but easier to use than the widely spread ultrametric distance.
Furthermore, we study several kinds of partitions of \( I \) which may be classical or "weakly" intersecting classes. Fuzzy partitions are considered and evaluated by means of the information they process: this quantity generalizes the classical Shannon's information of an ordinary partition, and it is increasing with respect to the refinement of partitions. We then compare two fuzzy partitions, by computing the gain of information obtained by replacing a fuzzy partition by another one. The following problems are addressed: aggregation of classes, allocation of an element to a class, selection, or approximation of fuzzy partition and evaluation of its accuracy.

I. METRICAL PROPERTIES OF FUZZY RELATIONS

Let \( E = \{x_1, \ldots, x_n\} \) a finite set of cardinality \( n \). A fuzzy binary relation \( R \) on \( E \) is a mapping from \( E \times E \) to \([0, 1] \subseteq \mathbb{R} \). We suppose that \( R \) is a \(*\)-relation \([3]\), i.e., \( R \) is

- definite: for \( x, y \) in \( E \), \( R(x, y) = 1 \) if and only if \( x = y \)
- symmetric: \( R(x, y) = R(y, x) \forall x, y \)
- \(*\)-transitive: for \( x, y \) in \( E \),

\[
R(x, y) \geq \bigvee_{z \in E} (R(x, z) \ast R(y, z)),
\]

for an operation \( \ast \) from \([0, 1] \times [0, 1] \) to \([0, 1] \), \( \vee \) denoting the supremum.

We further suppose that \( R \) verifies \( \forall x, y, z, \)

\[
R(x, z) \ast R(y, z) \geq ((R(x, z) + R(y, z) - 1) \vee 0).
\]

Then, \( d(x, y) = 1 - R(x, y) \) is a distance relation and satisfies the triangular inequality; \( d \) is the distance or T-distance associated with \( R \).

More particularly, we consider the following operations \( \ast \):

- if \( a \ast b = ab \), \( d \) is called a probabilistic distance or P-distance
- if \( a \ast b = a \wedge b \), \( d \) is an ultrametric distance, or UM-distance, \( \wedge \) denoting the infimum.

The following chain is easy to verify: \( d \) is UM-distance \( \Rightarrow \) \( d \) is P-distance \( \Rightarrow \) \( d \) is T-distance.

In Section 2, other chains will be studied between UM and T.
1. \( r \)-SPHERES AND \( r \)-CLIQUES

For a distance \( d \) and an element \( x \) of \( E \), the \( r \)-sphere of center \( x \) is defined by \( B_x(r) = \{ y \in E, d(x, y) \leq r \} \). A \( r \)-clique \( C(r) \) is a subset of \( E \) such that, for every \( x \) and \( y \) in \( C(r) \), \( d(x, y) \leq r \), and \( C(r) \) is maximal for inclusion. Obviously, for all \( x \) in \( C(r) \), \( B_x(r) \) contains \( C(r) \).

Let us characterize a UM-distance by means of \( r \)-spheres and \( r \)-cliques. A distance \( d \) is \( r \)-homogeneous if: \( B_x(r) \cap B_y(r) = \emptyset \) or \( B_x(r) = B_y(r) \) \( \forall x, y \) in \( E \). It can be proved that \( d \) is \( r \)-homogeneous if and only if, for every \( r \)-clique \( C(r) \) and any \( x \) in \( C(r) \), \( C(r) = B_x(r) \). We deduce that \( d \) is UM-distance if and only if \( d \) is \( r \)-homogeneous, for every \( r \in [0, 1] \).

Probabilistic distances possess the following property [3]: If \( d \) is a non-trivial P-distance, then \( d \) is \( r \)-homogeneous, for \( r = 1 - \bigwedge_{(x, y): R(x, y) \neq 0} R(x, y) \).

2. A FEW EXAMPLES OF \(*\)-RELATIONS

The most widely used (e.g., in cluster analysis) is associated to the ultrametric distance, previously defined. Some weaker relations are listed below:

\[
\begin{align*}
  a \ast_{\text{UM}} b &= a \wedge b & \text{(ultrametric distance)} \\
  a \ast_p b &= ab & \text{(probabilistic distance)} \\
  a \ast_{\text{m-SM}} b &= a \wedge b - (1 - a \vee b)/m & \text{\( m \)-supermetric distance)} \\
  a \ast_{\text{IP}} b &= (a \wedge b)(a + b - ab) & \text{(IP-distance)} \\
  a \ast_{\text{JP}} b &= ab(1 + a \wedge b)/(a + b) & \text{(JP-distance)}.
\end{align*}
\]

IP, JP, and \( m \)-SM-distances are introduced in [4], where we proved that \( m \)-supermetricity is a necessary condition for \( m \)-hypermetricity, defined as follows:

\[
\forall x_1, \ldots, x_m, y_1, \ldots, y_{m+1}, \text{ in } E,
\sum_{1 \leq i < j \leq m} d(x_i, x_j) \leq \sum_{1 \leq i < j \leq m+1} d(y_i, y_j) \leq \sum_{1 \leq i \leq m} \sum_{1 \leq j \leq m+1} d(x_i, y_j).
\]

The \( m \)-supermetricity, being defined from a \(*\)-relation, is much more convenient to use. In [4], we also prove that a IP-distance is 4-HM and a JP-distance is 2-HM.

We summarize the properties of the distances in Fig. 1, where \( A \rightarrow B \) means \( A \) implies \( B \), and \( A \cdot \cdot \cdot B \) means that \( A \) and \( B \) are incomparable.
3. Matricial Characterizations of \(\ast\)-Relations

For a relation \(R\), we denote \(M_R = (m_{ij})\) the \(n \times n\) matrix defined by
\[
m_{ij} = R(x_i, x_j), \quad 1 \leq i, j \leq n.
\]

**Definition 1.** For any binary operation \(\ast\) on \([0, 1]\), we define the \(\ast\)-composition of two \(n \times n\) matrices \(A = (a_{ij}), B = (b_{ij})\) by \(A \ast B = D\), with \(D = (d_{ij})\) such that
\[
d_{ij} = \bigvee_k (a_{ik} \ast b_{kj}).
\]

**Proposition 1.** Let \(M_R\) be a symmetric matrix with diagonal elements equal to 1, \(R\) is a \(\ast\)-relation if and only if
\[
M_R \ast M_R = M_R.
\]

**Proof.** If \(M_R \ast M_R = M_R\), then \(R(x_i, x_j) = \bigvee_k (R(x_i, x_k) \ast R(x_k, x_j))\) for all \(i, j\) and \(R\) is \(\ast\)-transitive.

Conversely, let \(M' = M_R \ast M_R\). If \(R\) is \(\ast\)-transitive, then, by definition
\[
m_{ij} \geq m'_{ij} - \bigvee_k (m_{ik} \ast m_{kj}) \quad \forall i, j.
\]

Now \(m'_{ij} \geq m_{ij} \ast m_{jj} = m_{ij} \ast 1\). As \(R\) is associated with a T-distance,
\[
R(x_i, x_j) \ast 1 \leq (R(x_i, x_j) \vee 0) = R(x_i, x_j).
\]
So
\[
m'_{ij} \leq m_{ij}, \quad \forall i, j,
\]
and
\[
M' = M_R.
\]

For \(d\) a UM-distance, the result is given in [8]. A simple charac-
terization follows for a distance $d$ in terms of its distance matrix $M_d = J - M_R$, where $J$ is the “all one” $n \times n$ matrix.

**Corollary.** A $n \times n$ matrix $M$ is realizable as a distance matrix iff $M$ is symmetric, has zero diagonal, and satisfies

$$(J - M)^* (J - M) = J - M$$

for some $*$. 

### 4. Partitions in Metric Spaces

#### 4.1. Crisp Partitions

The family $\mathcal{F} = \{F_1, \ldots, F_m\}$ of subsets of $E$ is a **crisp partition** if $F_i \subseteq E$, $\bigcup_i F_i = E$, and $F_i \cap F_j = \emptyset$ for $i \neq j$.

**Definition.** $\mathcal{F}$ will be called **homogeneous** for a distance $d$ if for all $x, y$ in $F_i$, $z$ in $F_j$, $i \neq j \Rightarrow d(x, y) < d(x, z)$.

$\mathcal{F}$ will be called **centered** for $d$ if for all $i$, there is a $c_i$ in $F_i$ (a center) such that $x \in F_i$ and $i \neq j \Rightarrow d(x, c_i) \leq d(x, c_j)$.

**Remark.** The center of a $F_i$ is not necessarily unique: if $F$ is homogeneous, then it is centered, every element of $F_i$ being a center. It is easy to prove the following.

**Proposition.** If $\mathcal{F}$ is a partition homogeneous for $d$, then $d$ is $r$-homogeneous with

$$r = \min_{i \neq j} d(F_i, F_j) = \min_{i, j} \min_{x \in F_i, y \in F_j} d(x, y).$$

Conversely, if $d$ is $r$-homogeneous, it yields homogeneous partitions by $r$-cliques.

#### 4.2. Fuzzy Partitions

We recall that a fuzzy partition of $E$ is a family $\Delta = \{E_i, 1 \leq i \leq m\}$ of fuzzy subsets $E_i$ of $E$ with membership function $\mu_i [1]$, such that

$$\sum_{x \in E} \mu_i(x) > 0 \quad \forall i \in \{1, \ldots, m\}$$

and

$$\sum_{i=1}^m \mu_i(x) = 1 \quad \forall x \in E.$$
A partition will be said spherical if for all \( i \) there exists a \( c_i \) in \( E \) such that \( \mu_i(x) = f(d(x, c_i)) \), with \( f: [0, 1] \to [0, 1] \). Modifying slightly a definition in [7], it will be called \( q \)-spherical if \( f \) is a function \( q \) such that \( q(a) = 0 \) and \( b > a \Rightarrow q(b) = 0 \).

A partition is uniform if \( |E_i| = k \ \forall i \in \{1, \ldots, m\} \), where \( |F| = \sum_{x \in F} \mu_F(x) \). Of course, if the space is regular, i.e., for all \( x \) in \( E \), \( |\{ y \in E, d(x, y) = a \}| \) depends only on \( a \), then all fuzzy spheres \( S(f, c_i) \) are equicardinal, and any spherical partition is uniform.

For a distance \( d \), we consider a family \( \mathcal{F} = \{F_i, 1 \leq i \leq m\} \) of \( r \)-cliques covering \( E \) (i.e. \( E \subseteq \bigcup_{i} F_i \)), with \( 0 < r < 1 \). \( \mathcal{F} \) induces a fuzzy partition \( \Delta = \{E_i, 1 \leq i \leq m\} \) of \( E \), every fuzzy subset \( E_i \) of \( E \) defined by one of the following membership functions:

\[
\mu_i(x) = \frac{\left( \sum_{y \in F_i} R(x, y) \right)^{1/m}}{\sum_{j=1}^{m} \sum_{y \in F_i} R(x, y)} \quad \forall x \in E, \quad (1)
\]

\[
\mu_i^{(2)}(x) = \begin{cases} 0 & \text{if } x \in F_i, \\ \frac{\left( \sum_{y \in F_i} R(x, y) \right)^{1/m}}{\sum_{j \in \mathcal{G}(x)} \sum_{y \in F_i} R(x, y)} & \text{if } x \notin F_i, \end{cases}
\]

where \( \mathcal{G}(x) = \{j \in \{1, \ldots, m\}, x \in F_j\} \) or

\[
\mu_i^{(3)}(x) = R(x, c_i) \left( \sum_{j=i}^{m} R(x, c_j) \right)^{-1/m} \quad \forall x \in E,
\]

in the case where \( \mathcal{F} \) is centered, with

\[
d(x, c_i) = d(x, c_j) \quad \forall x \in F_i \cap F_j, \quad i \neq j.
\]

From a partition \( \mathcal{F} \) of \( E \), we can deduce a fuzzy partition \( \Delta \), softening the belonging of every element to a class. In particular, if we suppose that

\[
\sum_{x \in E} R(x, y) = k, \quad \forall y \in E, \quad (4)
\]

a uniform partition \( \mathcal{F} \) yields a uniform fuzzy partition \( \Delta^{1} \) by using membership values \( \mu_i^{(1)}(x) \), for \( x \in E \).

**DEFINITION 3.** \( \Delta \) is an \( \eta \)-partition, for a threshold \( \eta \in [0, 1] \), if for all \( x \) in \( E \), there exists a unique \( i \) such that \( \mu_i(x) \leq \eta \).

**PROPOSITION 3.** If \( R \) verifies condition (4), a uniform partition \( \mathcal{F} \) of \( r \)-cliques of \( E \) yields an \( \eta \)-partition \( \Delta \), with \( \eta = a(1-r)/k \), \( a \) being the cardinality of any class of \( \mathcal{F} \).
Now, if we suppose that under condition (4) $\mathcal{F}$ is centered, we obviously deduce a spherical partition $\mathcal{A}^3$ by using membership values $\mu_i^{(3)}(x)$, for $x \in E$.

If $\mathcal{F}$ is a partition of $E$, the fuzzy partition $\mathcal{A}^2$ defined by membership functions $\mu_i^{(2)}$ is $\mathcal{F}$ itself.

4.3. Examples from Coding Theory

The general setting of coding theory is the following: We are given the set $E = (\mathbb{F}_2)^n$ of the binary sequences of length $n$ endowed with the Hamming distance $H(x, y) = |\{i, x_i \neq y_i\}|$ for $x = (x_i), y = (y_i)$ in $E$. An $e$-error correcting code is a subset $C$ of $E$ s.t. for any $c_1$ and $c_2$ in $C$, $H(c_1, c_2) \geq 2e + 1$, i.e., spheres of radius $e$ around codewords are disjoint. If furthermore these spheres partition $E$, $C$ is perfect. Thus a perfect code is equivalent to a spherical crisp partition of $E$. A code with a complete decoding algorithm is a centered partition, with codewords as centers: every received element of $E$ is clustered (or decoded) into a nearest codeword. From [5] can be deduced the following

**Proposition 4.** To every code $C = \{C_i\}$ can be associated a spherical fuzzy partition with the centers being the codeword and $\mu_j(x) = f(H(x, c_i))$, where $f: j \rightarrow x_j, j = 1, 2, ..., n$.

The $x_j$ are the coefficients of the expansion of the code characteristic polynomial in the basis of Krawtchouk polynomials. Here $f$ is from $\{1, ..., n\}$ to $\mathbb{Q}$. It is not known in general when $f$ is positive or is a $q$-function.

4.4. Quasi-Partitions

We now study special cases of coverings of $E$.

**Definition 4.** A distance $d$ is $k$-ultrametric ($k$-UM) if for any $k + 2$ points $x_1x_2 \cdots x_{k+2}$ in $E$, the $\binom{k+2}{2}$ distances between them satisfy

$$d(x_i, x_j) \leq \sqrt{\sum_{(i,j), (i,j) \neq \{i,j\}}}$$

The case $k = 1$ corresponds to $d$ a UM; a $k$-UM is also a $(k + 1)$-UM. It is equivalent to say that the distance $d$ is $k$-ultrametric if the maximum distance between any $k + 2$ points is realized at least twice.

We have proved in [3] that $k$-ultrametricity and weak $k$-ultrametricity, as defined in [8], are equivalent:

**Proposition 5.** A distance $d$ is $k$-ultrametric if and only if two distinct $r$-cliques intersect in at most $k - 1$ elements.
Let us consider the following generalization of partitions:

**Definition 5.** \( \mathcal{F} = \{F_i\} \) is a \( k \)-quasi-partition if \( E = \bigcup F_i \) and \( i \neq j \Rightarrow |F_i \cap F_j| \leq k - 1. \)

Of course, \( k = 1 \) gives the classical partitions and we have:

**Proposition 6.** \( d \) is \( k \)-ultrametric if and only if for any \( r \), the \( r \)-cliques form a quasi-partition of \( E \).

## 5. Some Practical Consequences

The properties of a given fuzzy relation \( R \) yield consequences which can have a practical interest. We give such particular results here.

### 5.1. \( m \)-SM-Distance

Let us consider an \( m \)-SM distance \( d \) and an \( r \)-clique \( C \), not containing an element \( y \) of \( E \). There exists \( x \in C \) such that \( d(x, y) > r \). For any other \( z \in C \) such that \( d(y, z) \geq d(x, z) \), we get

\[
r < d(x, y) \leq d(y, z) + \frac{1}{m} d(x, z) \leq d(y, z) + \frac{r}{m}
\]

and \( d(y, z) \geq ((m-1)/m)r \). This result is true for every \( z \) in \( C \) if \( d(x, y) \geq r(1+1/m) \). Now, for \( z \) and \( t \) in \( B_x(r) \), we have

\[
d(z, t) \leq \frac{m+1}{m} r
\]

and

\[
B_x(r) \subseteq C \left( \frac{m+1}{m} r \right).
\]

Consequently, \( B_x((mi(m+1))r) \subseteq C(r) \subseteq B_x(r) \).

We conclude that the \( m \)-SM condition, less restrictive than the UM-condition, separates \( E \) in such a way that:

- For a given \( r \)-clique \( C \) and a point \( y \) in \( C \), if there exists an \( x \) in \( C \) "far enough" from \( y \), namely \( d(x, y) \leq r(1+1/m) \), then \( y \) is "relatively far" from any element \( z \) of \( C \), namely \( d(y, z) \geq r(1-1/m) \).

- Any \( r \)-clique \( C(r) \) is between two \( r \)-spheres centered in any element of \( C(r) \). For large values of \( m \), \( C(r) \) may be assimilated to \( B_x(r) \). As the \( r \)-spheres are practically easier to determine, we almost partition \( E \) by \( r \)-spheres.
5.2. \( k \)-UM-Distances

Classical algorithms of cluster analysis tend to approach a given fuzzy relation \( R \) by another relation \( R' \) associated with a UM-distance. They imply the determination of \( (\frac{1}{2}) \) values of distances between two elements of \( E \).

If we approach \( R \) by a relation \( R'' \) associated with a \( k \)-UM-distance \( d'' \), the \( r \)-cliques defined by \( d'' \) intersect in at most \( k - 1 \) elements \([3]\). The number of distances to be determined is at most \( T(n, k + 1) + 1 \), for \( k \geq 2 \), where \( T(n, i) \) is the maximum number of edges in a graph with \( n \) vertices not containing a complete graph with \( i \) vertices (Turán number); it is for small \( k \) less than \( (\frac{1}{2}) \) \([4]\).

For \( n \) large, the number of elements belonging to two \( r \)-cliques, at most equal to \( k - 1 \), is very small with respect to \( n \), and the number of distances is approximately \( (\frac{k - 1}{k})(\frac{1}{2}) \). Thus, \( k \)-UM-distances might constitute a trade-off between complexity of computation and accuracy of clustering.

II. INFORMATION OF FUZZY PARTITIONS

We propose to study a criterium characterizing a fuzzy partition, measuring its fuzziness and enabling us to make a choice when facing several fuzzy partitions of a given set of events.

1. Associated Fuzzy and Crisp Partitions

Let \( \Delta = \{E_1, \ldots, E_m\} \) be a fuzzy partition defined on \( E = \{x_1, \ldots, x_n\} \). We denote by \( \mu_i \) the membership function defining every fuzzy class \( E_i \), and \( \mu_{ij} = \mu_i(x_j), 1 \leq i \leq m, 1 \leq j \leq n \).

For a given threshold \( \eta \in [0, 1] \), let \( \Delta^n \) be the set of crisp partitions of \( E \) corresponding to \( \Delta \): an element

\[
\delta = \{E_1^n, \ldots, E_m^n\}
\]

verifies

\[
x_j \in E_i^n \Rightarrow \mu_{ij} \geq \eta, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.
\]

If, for a value \( \eta \), there exists only a crisp partition in \( \Delta^n \), \( \Delta \) is an \( \eta \)-partition.

In the case where \( \Delta \) is a crisp partition, \( \Delta^n \) only contains the element \( \Delta \) for every value of \( \eta \).

In the sequel, we will denote by \( \Sigma \), a summation for the values of \( i \) between 1 and \( m \).
2. **Fuzzy Probability**

Let us now suppose that $p$ is a probability measure defined on $(E, \mathcal{P}(E))$, where $\mathcal{P}(E)$ denotes the set of crisp subsets of $E$. $F$ denotes the set of fuzzy subsets of $E$.

For any crisp partition $\delta = \{E_1^\gamma, ..., E_m^\gamma\} \in \Delta^n$, associated with $\Delta = \{E_1, ..., E_m\}$, we define the relative $\eta$-probability of $E_i$:

$$P_{\eta,\delta}(E_i) = \sum_{x_j \in E_i^\gamma} \mu_{ij} p(x_j) \quad 1 \leq i \leq m, \quad 1 \leq j \leq n$$

(5)

We easily verify the following results:

- If $E_i$ is a crisp subset of $E$, $P_{\eta,\delta}(E_i) = p(E_i^\gamma)$.
- $\forall i \in \{1, ..., m\}$, $p(E_i^\gamma) \geq P_{\eta,\delta}(E_i) \geq \eta p(E_i^\gamma)$.
- Let $\Delta' = \{E_1, ..., E_{m-2}, E_{m-1} \cup E_m\}$ and $\delta' = \{E_1^\gamma, ..., E_{m-2}^\gamma, E_m^\cap E_m^\gamma\}$.

Then

$$P_{\eta,\delta}(E_{m-1} \cup E_m) \geq P_{\eta,\delta}(E_{m-1}) + P_{\eta,\delta}(E_m)$$

$$P_{\eta,\delta}(E_i) = P_{\eta,\delta}(E_i) \quad \forall i \in \{1, ..., m-2\}.$$

- If $\Delta$ is a $\eta$-partition, $\delta$ is unique in $\Delta^n$; $P_{\eta,\delta}(E_i)$ is denoted by $P_\eta(E_i)$ and called the $\eta$-probability of $E_i$ [2]. Then, with $U_m = \bigcup_{i=1}^m E_i$, we have

$$P_\eta(U_m) = \sum_i P_\eta(E_i) \leq 1.$$  

(7)

The equality holds in (4) if and only if $U_m = E$. It is easy to see that $P_\eta$ constitutes a probability measure defined on $(E, \mathcal{F})$.

With respect to the $\eta$-probability, $\Delta$ may be considered as an incomplete system of fuzzy events, and any fuzzy classes $E_i$ and $E_j$ of $\Delta$ are incompatible with respect to $P_\eta$. Every class of the crisp partition of $\Delta^n$ is the $\eta$-level set of $E_i$, $1 \leq i \leq m$.

- Since $U_m$ is associated with the unique crisp partition $\{E\}$, we have, for every $\delta \in \Delta^n$,

$$\sum_i P_{\eta,\delta}(E_i) \leq P_\eta(U_m) \leq 1.$$  

(8)

The quantity $P_\eta(U_m)$ equals

$$\sum_{j=1}^n \left( \bigvee_{i=1}^m \mu_{ij} \right) p(x_j),$$

and lies in $[\eta, 1]$. 

409-116 1-12
• For another value \( \eta' \) of the threshold, with \( \eta' > \eta \), we obtain \( \Delta^{\eta'} \leq \Delta^{\eta} \) and \( P_{\eta'}(U_m) \geq P_{\eta}(U_m) \).

If \( \Delta \) is a \( \eta \)-partition, \( P_{\eta}(E_i) \geq P_{\eta'}(E_i), \forall i \in \{1, \ldots, m\} \).

Remark. If, for every \( x_j \in E \), there exists a grade of membership considerably greater than every other one, we can find \( \eta \) corresponding to an \( \eta \)-partition. In the opposite case, we can only define sets \( \Delta^{\eta} \) containing more than one element, for several values of \( \eta \) which are not very large.

3. Information Processed by a Fuzzy Partition

In order to treat of fuzzy partition in a way analogous to the classical study of crisp partitions, we define a measure of the information processed by \( \Delta \), with regard to an element \( \delta \) of \( \Delta^{\eta} \), for a given threshold \( \eta \).

**Definition 6.** The information processed by \( \Delta \) with regard to \( \delta \) is defined by

\[
I_{\eta,\delta}(\Delta) = -\left( \sum_i P_{\eta,\delta}(E_i) \log P_{\eta,\delta}(E_i)/P_{\eta}(U_m) \right).
\]

The \( \eta \)-information of \( \Delta \) is the quantity

\[
I_{\eta}(\Delta) = \bigvee_{\delta \in \Delta^{\eta}} I_{\eta,\delta}(\Delta).
\]

In the sequel, the logarithms will be taken to the base 2 and we will use the notation \( L(x) = -x \log x \).

If, for two values \( \eta \) and \( \eta' \), we have

\[
P_{\eta}(U_m) = P_{\eta'}(U_m)
\]

then

\[
I_{\eta'}(\Delta) \leq I_{\eta}(\Delta).
\]

In the case where \( \Delta \) is a \( \eta \)-partition, the \( \eta \)-information of \( \Delta \) [2] is defined by

\[
I_{\eta}(\Delta) = \sum_i L(P_{\eta}(E_i))/\sum_i P_{\eta}(E_i),
\]

and it has the classical form of Shannon's information expressed by Renyi [9] for an incomplete system of events.

Let us study the crisp partitions \( \delta \in \Delta^{\eta} \) which maximize \( I_{\eta,\delta}(\Delta) \), for a given threshold \( \eta \):
(a) If \( \eta > \frac{1}{2} \), \( A' \) contains at most one crisp partition \( \delta \); when \( \delta \) exists, the information processed by \( A \) with regard to \( \delta \) equals the \( \eta \)-information of \( A \).

(b) If \( \eta \leq \frac{1}{2} \), the elements of \( E \) belonging to two different classes \( E_i \) and \( E_k \) in two crisp partitions \( \delta \) and \( \delta' \) of \( A' \) correspond to grades of membership \( \mu_j \) and \( \mu_k \). Let \( x_j \) such an element, with \( y_j = \mu_j p(x_j) \), \( y_k = \mu_k p(x_j) \). Denote by \( q_i \) and \( q_k \) the relative \( \eta \)-probabilities \( P_{\eta,\delta}(E_i) \) and \( P_{\eta,\delta}(E_k) \).

**Proposition 7.** For \( \eta \geq \frac{1}{2} \), let \( x_j \) be an element having a smaller (resp. greater) grade of membership in \( E_i \) than in \( E_k \). The \( \eta \)-information of \( A \) is attained for a crisp partition \( \delta \in A' \) in which \( x_j \) belongs to the class \( E_i \) of smallest relative \( \eta \)-probability if \( P_{\eta,\delta}(E_i) \geq 1/e \) (resp. \( \leq 1/e \)).

**Proof.** The information processed by \( A \) with regard to \( \delta \) and \( \delta' \) are such that

\[
(I_{\eta,\delta}(A) - I_{\eta,\delta'}(A)) P_{\eta}(U_m) = L(q_i + y_i) + L(q_k) - L(q_i) - L(q_k + y_k).
\]

The function \( f(q) = L(q + y) - L(q) \) is decreasing for positive values of \( q \) and \( y \). Then, if \( q_i \leq q_k \), we get

\[
L(q_i + y_i) - L(q_i) \geq L(q_k + y_i) - L(q_k).
\]

Now, the function \( g(y) = L(q + y) - L(q) \) is decreasing for \( q + y \geq 1/e \). We obtain

\[
L(q_i + y_i) - L(q_i) \geq L(q_k + y_i) - L(q_k) \quad \text{and} \quad I_{\eta,\delta}(A) \geq I_{\eta,\delta'}(A)
\]

if

\[
y_i \geq y_k \quad \text{and} \quad P_{\eta,\delta}(E_i) \leq 1/e
\]

or if

\[
y_i \leq y_k \quad \text{and} \quad P_{\eta,\delta}(E_k) + y_i \geq 1/e.
\]

This last condition is realized if \( P_{\eta,\delta}(E_i) \geq 1/e \).

(c) If \( \eta = \frac{1}{2} \), the elements of \( E \) belonging to two different classes \( E_i \) and \( E_k \) in two crisp partitions \( \delta \) and \( \delta' \) of \( A' \) correspond to a grade of membership equal to \( \frac{1}{2} \); thus:

**Corollary.** For \( \eta = \frac{1}{2} \), the \( \eta \)-information of \( A \) is attained for a partition \( \delta \in A' \) in which an element \( x_j \) such that \( \mu_j = \mu_k = \frac{1}{2} \) belongs to the class \( E_i \) of smallest relative \( \eta \)-probability.
EXAMPLE. Consider \( n = 4, m = 3, \eta = \frac{1}{2}, \mathcal{A}^n = \{ \delta, \delta' \} \),

\[
\begin{array}{c|c|c|c}
| E_1 & E_2 & E_3 |
\hline
0.7 & 0.2 & 0.1 \\
\hline
0.5 & 0.5 & 0 \\
\hline
0.3 & 0.6 & 0.1 \\
\hline
0.3 & 0.1 & 0.8 \\
\end{array}
\quad
\begin{array}{c|c|c|c}
| x_1 & 1 & 0 & 0 \\
\hline
| x_2 & 1 & 0 & 0 \\
\hline
| x_3 & 0 & 1 & 0 \\
\hline
| x_4 & 0 & 0 & 1 \\
\end{array}
\quad
\begin{array}{c|c|c|c}
| \delta & 1 & 0 & 0 \\
\hline
| \delta' & 0 & 1 & 0 \\
\hline
| & 0 & 1 & 0 \\
\hline
| & 0 & 0 & 1 \\
\end{array}
\]

\[ I_\eta(\mathcal{A}) = I_{\eta, \delta}(\mathcal{A}) \geq I_{\eta, \delta'}(\mathcal{A}) \quad \text{if} \quad p(x_1) < p(x_3), \quad (i) \]
\[ = I_{\eta, \delta}(\mathcal{A}) \geq I_{\eta, \delta'}(\mathcal{A}) \quad \text{if} \quad p(x_1) \geq p(x_3). \quad (ii) \]

The element \( x_3 \) will be in class \( E_1^{\eta} \) in the case (i) and in class \( E_1^{\eta} \) in case (ii), where we maximize the information processed by \( \mathcal{A} \).

4. Refinement of Partitions

In order to use the information processed by a fuzzy partition in aggregating methods, let us look for results generalizing the branching property of Shannon's information.

For a given fuzzy partition \( \mathcal{A} \), let us consider the crisp partition \( \delta \in \mathcal{A}^n \). We group the elements of \( E_{m - 1} \) and \( E_m \) in a new fuzzy class \( E_0 = E_{m - 1} \cup E_m \); thus, we define a fuzzy partition \( \mathcal{A}' \) and an associated crisp partition \( \delta' \) as in (3). The union of the classes of \( \mathcal{A} \) and \( \mathcal{A}' \) is the same \( U_m \in \mathcal{F} \).

PROPOSITION 8. The refinement of a fuzzy partition \( \mathcal{A}' \), when we consider the corresponding refinement of a crisp partition \( \delta' \) of \( \mathcal{A}^n \) increases the information of \( \mathcal{A}' \) with regard to \( \delta' \), if the relative \( \eta \)-probabilities of the two classes deduced from the refined class \( E_0 \) either is at least equal to 1/e or has a sum equal to the relative \( \eta \)-probability of \( E_0 \).

Proof. We obtain

\[
(I_{\eta, \delta}(\mathcal{A}) - I_{\eta, \delta'}(\mathcal{A})) P_\eta(U_m) = L(P_{\eta, \delta}(E_{m - 1})) + L(P_{\eta, \delta}(E_m)) - L(P_{\eta, \delta'}(E_0)).
\]

As \( P_{\eta, \delta}(E_0) \geq \sum_{i = m - 1, m} P_{\eta, \delta}(E_i) = p \), the right-hand side of this equality is

\[
p \left( \sum_{i = m - 1, m} L \left( \frac{1}{p} P_{\eta, \delta}(E_i) \right) + L(p) - L(P_{\eta, \delta}(E_0)) \right). \quad (11)
\]
PARTITIONS AND FUZZINESS

- If \( P_{\eta,\delta}(E_0) = \rho \), and particularly if \( A \) is an \( \eta \)-partition, relation (11) gives

\[
I_{\eta,\delta}(A) - I_{\eta,\delta}(A') = \frac{\rho}{P_{\eta}(U_m)} \sum_{i=1, m} L\left(\frac{1}{\rho} P_{\eta,\delta}(E_i)\right)
\]

which is a kind of conditional information processed by \( E_{m-1} \) and \( E_m \) for a given fuzzy subset \( E_0 \) of \( E \). Consequently, it is positive and relation (12) generalizes the classical branching property:

- If \( P_{\eta,\delta}(E_0) > \rho \geq 1/\epsilon \), the left-hand side of (12) is not less than the right-hand side and thus, it is also positive.
- If \( \rho < P_{\eta,\delta}(E_0) \leq 1/\epsilon \), the inequality above-mentioned holds in the opposite sense and \( I_{\eta,\delta}(A) \) may be less or greater than \( I_{\eta,\delta}(A') \).

**COROLLARY.** The refinement of a fuzzy \( \eta \)-partition \( A \), when we consider the corresponding refinement of the associated crisp partition \( \delta \), increases the \( \eta \)-information.

As \( A'' \) contains all the \( \delta \) deduced from \( \delta' \in A'' \), we conclude that

\[
\bigvee_{\delta \in A''} I_{\eta,\delta}(A) \geq \bigvee_{\delta' \in A''} I_{\eta,\delta}(A) \geq \bigvee_{\delta' \in A''} I_{\eta,\delta}(A'),
\]

which means that

\[
I_{\eta}(A) \geq I_{\eta}(A').
\]

**PROPOSITION 9.** Under the same condition as in Proposition 8, the refinement of a fuzzy partition increases its \( \eta \)-information.

5. Sharpening of Fuzzy Classes

We would like to establish a difference between a partition in which the elements belong with a low grade of membership to all the classes, and a partition in which every element corresponds to one class with a prevailing grade of membership.

**DEFINITION 7.** Let \( A \) and \( A' \) be two fuzzy subsets of \( E \), with membership functions \( \mu_A \) and \( \mu_{A'} \). Then, \( A' \) is \( \eta \)-sharper than \( A \) if

\[
\mu_{A'}(x) \geq \mu_A(x) \quad \forall x \text{ such that } \mu_A(x) \geq \eta,
\]

\[
\mu_{A'}(x) < \eta \quad \forall \text{ other } x.
\]

Let \( A = \{E_1, \ldots, E_m\} \) and \( \{E'_1, \ldots, E'_m\} \) be two \( \eta \)-partitions of \( E \) such that \( E'_i \) is \( \eta \)-sharper than \( E_i \) and \( P_\eta(E'_j) = P_\eta(E_j) \forall j \neq i \).
PROPOSITION 10. If, for any \( j \neq i \), \( P_{\eta}(E_i) \leq P_{\eta}(E'_i) \leq (1/e) P_{\eta}(E_j) \), then \( I_{\eta}(A') \geq I_{\eta}(A) \).

Proof. Let \( x = P_{\eta}(E_i) \), \( k_1 = \sum_{j \neq i} L(P_{\eta}(E_j)) \), \( k_2 = \sum_{j \neq i} P_{\eta}(E_j) \),

\[
I_{\eta}(A) = (L(x) + k_1)/(x + k_2) = f(x).
\]

The derivative \( f'(x) \) of \( f \) is positive if and only if

\[-x - k_2 \log x - k_2 - k_1 \geq 0\]

or

\[
\sum_j P_{\eta}(E_j) \left( 1 + \log \frac{x}{P_{\eta}(E_j)} \right) \leq 0,
\]

which is realized if \( x \leq (1/e)P_{\eta}(E_i) \), \( \forall j \neq i \). Then, Proposition 10 is proved.

We remark that the criterium of information does not solve the problem exposed at the beginning of this section for every fuzzy partition. We complete this study in the next section.

6. Comparison of Fuzzy Partitions

In numerous practical applications, we must compare partitions of a given set \( E \). In the case of fuzzy partitions \( A = \{E_i, 1 \leq i \leq m\} \) and \( A' = \{E'_i, 1 \leq i \leq m\} \), with \( U_m = \bigcup_{i=1}^m E_i \), \( U'_m = \bigcup_{i=1}^m E'_i \) such that \( P_{\eta}(U_m) \leq P_{\eta}(U'_m) \), we define the following quantity, for a given threshold \( \eta \):

DEFINITION 8. The gain of information resulting from the replacement of \( A \) by \( A' \), with regard to crisp partitions \( \delta \in A^n \) and \( \Sigma \in A' \) is

\[
H_{n,\delta,\varepsilon}(A' \parallel A) = \left( \sum_i P_{n,\varepsilon}(E'_i) \log(P_{n,\varepsilon}(E'_i)/P_{n,\delta}(E'_i)) \right)/P_{\eta}(U'_m).
\]

It is easy to verify that this quantity is positive and null if and only if:

\[
P_{n,\varepsilon}(E'_i) = P_{n,\delta}(E'_i) \forall i.
\]

DEFINITION 9. The \( \eta \) gain of information resulting from the replacement of \( A \) by \( A' \) is defined as:

\[
H_{\eta}(A' \parallel A) = \bigwedge_{\delta \in A^n} H_{n,\delta,\varepsilon}(A' \parallel A).
\]

This quantity measures the nearness of \( A \) and \( A' \).
For two values $\eta$ and $\eta'$ such that $P_\eta(U_m') = P_{\eta'}(U_m')$, the inequality $\eta \leq \eta'$ implies that:

$$H_\eta(A' \parallel A) \geq H_{\eta'}(A' \parallel A).$$

In the case where $A$ and $A'$ are $\eta$-partitions, we obtain [2] $H_\eta(A' \parallel A) = (\sum_i P_\eta(E_i') \log(P_\eta(E_i')/P_\eta(E_i))) / \sum_i P_\eta(E_i).$

**Definition 10.** On the set $E$, the fuzzy partition $A'$ is $\eta$-sharper than the fuzzy partition $A$ if $E_i'$ is $\eta$-sharper than $E_i$ for every $i \in \{1, \ldots, m\}$.

If $A$ is $\eta$-sharper than $A'' = \{E_i'', 1 \leq i \leq m\}$, the sets $A''$ and $A'''$ are identical. Thus, for every $\delta \in A''$ and every $E_i \in \delta$, we have

$$P_{\eta,\delta}(E_i') \leq P_{\eta,\delta}(E_i)$$

and the inequality

$$H_{\eta,\delta}(A' \parallel A'') \geq H_{\eta,\delta}(A' \parallel A) \quad \forall \delta \in A''',$

implies

$$H_\eta(A' \parallel A'') \geq H_\eta(A' \parallel A).$$

This yields the following result:

**Proposition 11.** Let $A$, $A'$, and $A''$ be three fuzzy partitions of $E$. If $A$ is $\eta$-sharper than $A''$, then it produces a smaller $\eta$-gain of information than $A''$ when replaced by $A'$.

Under the condition $P_\eta(U_m) \leq P_{\eta'}(U_m)$, this property means that the sharper a fuzzy partition $A$ is, the closer to $A'$ it will be.

For a given crisp partition $\delta$, the sharpest fuzzy partition $A$ such that $\delta \in A''$ is $\delta$ itself.

In the particular case where $A' = \delta = \{E_i', 1 \leq i \leq m\} \in A''$, the gain of information between $A$ and $\delta$ admits interesting properties. We introduce the notation

$$H_\eta(\delta; A) = H_{\eta,\delta}(\delta \parallel A) = \sum_i p(E_i') \log(p(E_i')/P_{\eta,\delta}(E_i)).$$

and we remark that

$$H_\eta(\delta; A) \geq I_{\eta,\delta}(A) - I(\delta),$$

where $I(\delta)$ is the Shannon's information of $\delta$.

The gain of information $H_\eta(\delta; A)$ measures the fuzziness included in $A$,.
for a given crisp partition $\delta$. Its maximum value corresponds to $A$ defined by membership functions verifying

$$\forall i \in \{1, \ldots, m\} \mu_{ij} = \eta \quad \forall e_j \in E_i^\eta.$$  

Let us consider $A'$ defined by grades of membership $\mu'_{ij}$, $1 \leq i \leq m$, $1 \leq j \leq n$, and also corresponding to $\delta \in A''$. It is spreader than $A$ if

$$\forall i \in \{1, \ldots, m\} \forall x_j \in E_i^\eta, \mu'_{ij} = \bigvee_{k=1}^{m} \mu_{kj} \leq \bigvee_{k=1}^{m} \mu_{kj} = \mu_{ij}.$$  

It is easy to see that $A'$ is spreader than $A$ if $A$ is sharper than $A'$, but this condition is not necessary. We deduce the inequality

$$H_\eta(\delta, A') \geq H_\eta(\delta; A).$$

**Proposition 12.** For a given crisp partition $\delta$, the gain of information $H_\eta(\delta; A)$ is an increasing function of the spread of $A$.

### III. Conclusion

Crisp partitions are often used in classical spaces, especially in pattern recognition and classification. The maximization of Shannon's information processed by a partition under certain constraints, or the minimization of Kullback's gain of information (or divergence) obtained by replacing a partition by another one, are very useful tools.

When working with fuzzy partitions, we need similar concepts, the $\eta$-information and the $\eta$-gain of information above studied provide such tools.

For a given fuzzy partition $A$, how to choose an associated crisp partition $\delta$ is a problem which may be solved by using the $\eta$-information of $A$.

Between two different fuzzy partitions, which one is closer to another given fuzzy partition $A$? We propose to choose the one which minimizes the $\eta$-gain of information with respect to $A$.

To aggregate two classes of a fuzzy partition and increase the information as much as possible, we maximize the quantity (11).

If an element belongs to two fuzzy classes with a grade of membership equal to $\frac{1}{2}$, with which fuzzy class are we going to associate it? We will use the criterion of $\eta$-information to come to a decision.

To evaluate the spread or the accuracy of a fuzzy partition $A$, we look for the value of the $\eta$-gain of information of $A$ with respect to any associated crisp partition $\delta$. Numerous other decision problems will be solved by using $\eta$-information and $\eta$-gain, in practical applications.
REFERENCES


