Prépublication ou Rapport de Recherche
du
Departement de Mathématiques et Informatique
****

UNIVERSITE DES ANTILLES ET DE LA GUYANE
FACULTE DES SCIENCES EXACTES ET NATURELLES
Campus de Fouillole 97159 Pointe-à-Pitre (GUADELOUPE)

Fax 19 (590) 93 86 98 - Tel 19 (590) 93 86 93

N° : 95/07

ARCONTE Alain
KHATTABI Abdelaouahad
MANUCEAU Jérome
MARTIAS Claude

MUTUAL INFORMATION BETWEEN
Σ -ALGEBRAS

Réalisé avec le soutien de :
- l'Association des Amis de la Bibliothèque Universitaire
- l'Association des Mathématiciens et Informaticiens des Antilles et de la Guyane.

MUTUAL INFORMATION BETWEEN Σ-ALGEBRAS

A.ARCONTE, A.KHATTABI, J.MANUCEAU, C. MARTIAS
Universite des ANTILLES et de la GUYANE
Faculté des sciences Exactes et Naturelles
Département de Mathématiques-Informatique
971159 POINTE A PITRE -Cédex

ABSTRACT:

The well known notions of conditional entropy (in SCHANNON's sens) and more generally mutual conditional information between finite partitions are generalized to any σ-algebras and we study their properties. If $\mathfrak{C}$ is an any σ-algebra we identify the set $\mathcal{E}_1(\mathfrak{C})$ of finite conditional entropy knowing the σ-algebra $\mathfrak{C}$, and prove that it is a complet metric space.

Lastly we give a relation between the entropy defined upper and the differential entropy (*resp. entropy of a sequence of random variables*)

INTRODUCTION

Notions of entropy, mutual information and, more generally, conditional mutual information between finite random variables (or finite partitions) are well-known in Information Theory and have been successfully applied in various domains (see for example [1], [3], [4], [5], [8], [9] and the references in).

We develop the notion of conditional mutual information, (resp. conditional entropy) between any σ-algebras.

The first section is devoted to notations and definitions. We introduce in the second-one the notions mentioned above for the case of finite-type σ-algebras, i.e., completed and generated by a finite number of non-null events. The generalization to any σ-algebra is done in a natural way in the last section.

I) PRELIMINARIES

I-1) NOTATIONS AND DEFINITIONS

*) Let $(\Omega, \mathcal{F}, P)$ be a complete probability space ; we denote by $\sigma(\mathcal{A})$, $\mathcal{A}$ being a subset of $\mathcal{F}$, (resp. $\sigma'(\mathcal{A})$) the σ-algebra (resp. the complete σ-algebra) generated by $\mathcal{A}$.

*) An atom is an element A of $\mathcal{F}$ such that : $P(A) \neq 0$ and, for all $B \in \mathcal{F}$ such that $B \subset A$, we have $P(B) = P(A)$ or $P(B) = 0$.

REMARKS :

•) If A_1 and A_2 are two different atoms, then $P(A_1 \cap A_2)=0$.

•) The set of all atoms is at most countable.

*) A (finite or infinite) sequence $(A_n)_n$ of elements in $\mathcal{F}$ is a partition of the probability space $(\Omega, \mathcal{F}, P)$ if it is a (algebraic) partition of the set Ω with non-null events.

*) If $(A_n)_n$ is a finite (resp. countable) partition of the probability space, $\sigma((A_n)_n)$ is a finite (resp. countable) σ-algebra ; then, $\sigma'((A_n)_n)$ is called finite-type (resp. countable-type) σ-algebra. The elements A_n are the atoms of the following probability space:

$$(\Omega, \sigma'((A_n)_n), P_{|\sigma'((A_n)_n)})$$

*) The set of finite-type σ-algebras is denoted by $\mathcal{E}_o$.

REMARKS :

*) If a complete σ-algebra $\mathcal{A}$ is generated by a countable set $(A_n)_n$ of non-null events, then there exists an increasing sequence $(\mathcal{A}_n)_n$ of finite σ-algebras such that : $\mathcal{A} = \sigma'(\bigcup_n \mathcal{A}_n)$.

In fact, take $\mathcal{A}_n = \sigma(A_1, \ldots, A_n)$. The inverse is trivial. Such σ-algebras are called separable-type σ-algebras.

*) A separable type σ-algebra is not in general a countable-type σ-algebra. In fact, consider the complete probability space ($[0,1]$, $\sigma'(]\alpha_i, \beta_i[, i \in \mathbb{N}), \lambda$) where, for each integer i, α_i and β_i are elements of $\mathbb{Q} \cap [0,1]$ and λ is the Lebesgue measure on $[0,1]$; the σ-algebra:
$\sigma'((]\alpha_i, \beta_i[), i \in \mathbb{N})$ is a separable-type one but not a countable-type one ; in fact, no atoms exist in the considered space.

*) Usualy, a separable σ-algebra is, by definition, generated by a countable family of events which can be negligeable (see for example [6]). In the above example, the σ-algebra is a separable type σ-algebra in the sense defined in this paper but not in the usual sense.

*) Let $\mathcal{A} = \sigma'(\bigcup_n \mathcal{A}_n)$ be a separable-type σ-algebra, $(\mathcal{A}_n)_n$ being an increasing sequence of finite σ-algebras. If, for each n, we denote by $(\Xi_n) = (A_n^1, \ldots. A_n^n)$ the partition which generates $\mathcal{A}_n$, we always suppose all along this paper that (Ξ_{n+1}) is finest than (Ξ_n).

I-2) THEOREM :

(i) Any separable-type σ-algebra can be identified to a complete separable metric space.

(ii) Any sub-σ-algebra of a separable-type σ-algebra is a separable-type one.

PROOF :

(ii) is a consequence of (i) ; we show (i).

Let $\mathcal{F}$ to be a separable-type σ-algebra on Ω, endowed with a probability P. It is well known (see [7] for example) that the application:

$$(1)\ d : (A,B) \in F \times F \longrightarrow d(A,B) = P(A \,\Delta\, B) \in \mathbb{R}^+$$

defines a distance on the set $\mathcal{F}/\mathcal{N}$, where we denote by $\mathcal{N}$ the set of all null events.

Let $(A_n)_n$ to be a Cauchy sequence such that:

$$(2)\ d(A_n, A_{n+1}) \le \frac{1}{2^n},$$

We show in the following that $\lim A_n = \lim_n \sup A_n = \bigcap_n \bigcup_{p=n}^{\infty} A_p$.

Put $B_n = \bigcup_{p=n}^{\infty} A_p$; (B_n) being a decreasing sequence, it is easy to see that the limit of this sequence in the sense of the above distance defined on $\mathcal{F}$, is $\lim\sup A_n$. Then, it suffices to have : $d(A_n, B_n) \longrightarrow 0$, $(n \longrightarrow \infty)$.

Now, (1) $\Rightarrow$ $d(\bigcup_{p=0}^{k+1} A_{n+p}, \bigcup_{q=0}^{k} A_{n+q}) \leq \frac{1}{2^{n+k}}$.

Then, $d(\bigcup_{p=0}^{k+1} A_{n+p}, A_n) \leq \frac{1}{2^{n-1}}$,

and, $\lim_{k \longrightarrow \infty} d(\bigcup_{p=0}^{k} A_{n+p}, A_n) = d(B_n, A_n) \leq \frac{1}{2^{n-1}}$.

REMARKS :

(1) Let $\mathcal{A}$ be an algebra ; then $\overline{\mathcal{A}}$, the closure of $\mathcal{A}$ in the sense of the distance (1), is the σ-algebra $\sigma'(\mathcal{A})$. In fact, $\forall A \in \sigma'(\mathcal{A})$, $\forall \varepsilon > 0$, $\exists A' \in \mathcal{A}$ such that : $P(A \Delta A') < \varepsilon$.

(2) The notion of separable-type σ-algebra that we have been considering coincides with the notion of separability on complete metric space.

I-3 PROPOSITION

Suppose $\mathfrak{F}$ to be a separable-type σ-algebra. Then, x being any element of Ω :

-) either x belongs to an atom,

-) either there exists a decreasing sequence $(B_n)_n \subset \mathcal{F}$ which contains x, such that, for each n, $P(B_n) \neq 0$ and $\lim_n P(B_n) = 0$.

PROOF :

$\mathcal{F} = \sigma'(\bigcup_n \mathcal{A}_n)$, $(\mathcal{A}_n)$ being an increasing sequence of finite σ-algebras. For each n, let us denote by $A_n^{i(x)}$ the element of the partition in $\mathcal{A}_n$ wich contains x.
$(A_n^{i(x)})_n$ is a decreasing sequence in $\mathcal{F}$, each element of this partition containing x, and $P(A_n^{i(x)}) \neq 0$. Either $\bigcap_n A_n^{i(x)}$ is an atom containing x, either $(A_n^{i(x)})_n$ is the desired sequence.

II) INFORMATION, ENTROPY OF FINITE-TYPE Σ-ALGEBRAS.

II-1) DEFINITIONS

$(A_1, A_2, \ldots, A_n)$, $(B_1, B_2, \ldots, B_m)$ are finite partitions which generate respectively the elements $\mathcal{A}$, and $\mathcal{B}$ of $\mathcal{E}_0$:

$\mathcal{C}$ is an any σ-algebra,

$\mathfrak{z}$ is the continue function given by

$$\mathfrak{z}: [0,1] \longrightarrow [0,1] \quad \mathfrak{z}(x) = -x\mathrm{Log}(x) \text{ if } x \in]0,1]$$
$$\mathfrak{z}(0) = 0.$$

We define:

i)- the mutual information function between $\mathcal{A}$ and $\mathcal{B}$ given $\mathcal{C}$:

$$\mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{C}) = \sum_{i,j} P(A_i \cap B_j|\mathcal{C}).\mathrm{Log}\ \frac{P(A_i \cap B_j|\mathcal{C})}{P(A_i|\mathcal{C}).P(B_j|\mathcal{C})} .$$

- The mutual information between $\mathcal{A}$ and $\mathcal{B}$ given $\mathcal{C}$:

$$I(\mathcal{A},\mathcal{B}|\mathcal{C}) = E\left[\mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{C})\right]$$

ii)-the entropy function of $\mathcal{A}$ given $\mathcal{C}$, the non-negative function:

$$\mathcal{H}(\mathcal{A}|\mathcal{C}) = \mathcal{I}(\mathcal{A},\mathcal{A}|\mathcal{C}) = \sum_i \mathfrak{z}[P(A_i|\mathcal{C})]$$

- the entropy of $\mathcal{A}$ given $\mathcal{C}$:

$$H(\mathcal{A}|\mathcal{C}) = E\left[\mathcal{H}(\mathcal{A}|\mathcal{C})\right]$$

REMARKS :

-) The function $\mathcal{I}$ and $\mathcal{H}$ do not depend on the choice of partitions generating $\mathcal{A}$ and $\mathcal{B}$ respectively.

-) $\mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{C}) = \mathcal{H}(\mathcal{A}|\mathcal{C}) + \mathcal{H}(\mathcal{B}|\mathcal{C}) - \mathcal{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C})$ $a.a.$

-) Let $\mathcal{F}_0$ be the σ-algebra generated by the null events and $\mathcal{C}$ one element of $\mathcal{E}_0$ generated by $(C_1, C_2, \ldots, C_p)$, we fined the formulars of the usuals information I and entropy H by:

$$*) I(\mathcal{A},\mathcal{B}|\mathcal{C}) = E\left[\mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{C})\right]$$

$$= \sum_{i,j,k} P(A_i \cap B_j \cap C_k).\mathrm{Log}\ \frac{P(A_i \cap B_j|C_k)}{P(A_i|C_k).P(B_j|C_k)}$$

**) $H(\mathcal{A}|\mathcal{C}) = E\Big[\mathcal{H}(\mathcal{A}|\mathcal{C})\Big] = \sum_{i,k} P(C_k).\ \mathfrak{z}[P(A_i|C_k)]$

$= H(\mathcal{A} \vee \mathcal{C}) - H(\mathcal{C}).$

and $H(\mathcal{A}) = H(\mathcal{A}|\mathcal{F}_0) = \mathcal{H}(\mathcal{A}|\mathcal{F}_0) = \sum_i \mathfrak{z}[P(A_i)].$

***) $I(\mathcal{A},\mathcal{B}) = I(\mathcal{A},\mathcal{B}|\mathcal{F}_0) = \mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{F}_0)$

$$= \sum_{i,j} P(A_i \cap B_j)\ \text{Log}\ \frac{P(A_i \cap B_j)}{P(A_i).P(B_j)}.$$

$= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A} \vee \mathcal{B})$

$= H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}).$

II-2) PROPERTIES OF $\mathfrak{z}$.

1) $\mathfrak{z}$ is non-negative and strictly concave.

2) $\mathfrak{z}$ is sub-additive and then σ-sub-additive

Proof:
The sub-additivity of $\mathfrak{z}$ comes from the fact that the function t define by $t(x) = \mathfrak{z}(x) + \mathfrak{z}(y) - \mathfrak{z}(x+y)$ where, $y \in [0,1]$ and $x \in [0, 1-y]$ has a non-negative derivative and that $t(0) = 0$. It is easy to prove the others properties.

II-3) CONDITIONAL INDEPENDANCE OF Σ-ALGEBRAS.

We say that $\mathcal{A}$ and $\mathcal{B}$ are conditionally independant given $\mathcal{C}$, and we denote $\mathcal{A} \perp\!\!\!\perp_{\mathcal{C}} \mathcal{B}$ if:

$\forall (A,B) \in \mathcal{A} \times \mathcal{B},\ P(A \cap B|\mathcal{C}) = P(A|\mathcal{C}).P(B|\mathcal{C})\ a.s.$

PROPOSITION.

$\forall \mathcal{A},\mathcal{B} \in \mathcal{E}_0 \quad \forall\ \mathcal{C}\ \sigma$-algebra

1) $\mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{C}) \geq 0$ *i.e* $\mathcal{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) \leq \mathcal{H}(\mathcal{A}|\mathcal{C}) + \mathcal{H}(\mathcal{B}|\mathcal{C})$

2) $\mathcal{I}(\mathcal{A},\mathcal{B}|\mathcal{C}) = 0\ a.s \Leftrightarrow \mathcal{A} \perp\!\!\!\perp_{\mathcal{C}} \mathcal{B} \Leftrightarrow I(\mathcal{A},\mathcal{B}|\mathcal{C}) = 0$

Proof:

$$\mathcal{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) - \mathcal{H}(\mathcal{B}|\mathcal{C}) = \sum_{i,j} P(B_j|\mathcal{C}).\mathfrak{z}\Big[\frac{P(A_i \cap B_j|\mathcal{C})}{P(B_j|\mathcal{C})}\Big]$$

Jensen's inegality proves (1) and (2) is a consequence of the strictly concavity of the function $\mathfrak{z}$.

II-4) FUNDAMENTAL PROPERTIES.

$\mathcal{A}$ and $\mathcal{B}$, are elements of $\mathcal{E}_0$, $\mathcal{C}$ and $\mathcal{C}'$ are any σ-algebras

i) $\mathcal{B} \subset \mathcal{C} \iff \mathcal{H}(\mathcal{B}|\mathcal{C}) = 0$

and if $\mathcal{C} \in \mathcal{E}_0$ $\quad \mathcal{C} \subset \mathcal{B} \iff \mathcal{H}(\mathcal{B}|\mathcal{C}) = \mathcal{H}(\mathcal{B}) - \mathcal{H}(\mathcal{C})$

ii) $\mathcal{A} \subset \mathcal{B} \Rightarrow \mathcal{H}(\mathcal{A}|\mathcal{C}) \leq \mathcal{H}(\mathcal{B}|\mathcal{C})$

iii) $\mathcal{C}' \subset \mathcal{C} \Rightarrow H(\mathcal{A}|\mathcal{C}) \leq H(\mathcal{A}|\mathcal{C}')$

iv) $\mathcal{B} \subset \mathcal{C} \Rightarrow H(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) = H(\mathcal{A}|\mathcal{C})$ and $\mathcal{A} \underset{\mathcal{C}}{\perp\!\!\!\perp} \mathcal{B}$

Proof: It is easy to prove (ii), (iv) and the necessary condition of i).

For the sufficient condition of i), let (B_j) a finite partition which generate $\mathcal{B}$; we have $\mathcal{H}(\mathcal{B}|\mathcal{C}) = \sum_j \mathfrak{z}\left[P(B_j|\mathcal{C})\right]$.

If $\mathcal{H}(\mathcal{B}|\mathcal{C}) = 0$, $\sum_j \mathfrak{z}\left[P(B_j|\mathcal{C})\right] = 0$ and $\forall j$, $\mathfrak{z}\left[P(B_j|\mathcal{C})\right] = 0$

a.s then $P(B_j|\mathcal{C})$ is a characteristic function a. s.; so that,

$E\left[1_{B_j}|\mathcal{C}\right] = 1_C$ with $C \subset \mathcal{C}$.

From $E\left[(1_{B_j} - 1_C).1_C\right] = 0$ we deduce that $P(B_j \cap C) = P(C)$ and $C \subset B_j$ a. s.

But $E\left[1_{B_j}|\mathcal{C}\right] = 1_C \Rightarrow P(B_j) = P(C)$. Then $B_j = C$ a. s and $\mathcal{B} \subset \mathcal{C}$.

For iii) we have, $\mathcal{H}(\mathcal{A}|\mathcal{C}) = \sum_i \mathfrak{z}[P(A_i|\mathcal{C})]$ and:

$$E\left[\mathcal{H}(\mathcal{A}|\mathcal{C})|\mathcal{C}'\right] = \sum_i E\left(\mathfrak{z}\left[P((A_i|\mathcal{C})\right]|\mathcal{C}')\right) \leq \sum_i \mathfrak{z}\left[\ E(P(A_i|\mathcal{C})|\mathcal{C}')\right]$$

But $\quad E\left[P(A_i|\mathcal{C})|\mathcal{C}'\right] = P(A_i|\mathcal{C}')$

so that $\quad E\left[\mathcal{H}(\mathcal{A}|\mathcal{C})|\mathcal{C}'\right] \leq \mathcal{H}(\mathcal{A}|\mathcal{C}')$, we can then deduce that:

$E\left\{\ E\left[\mathcal{H}(\mathcal{A}|\mathcal{C})|\mathcal{C}'\right]\ \right\} \leq E\ \left\{\mathcal{H}(\mathcal{A}|\mathcal{C}')\right\}$ and $H(\mathcal{A}|\mathcal{C}) \leq H(\mathcal{A}|\mathcal{C}')$.

II-5) CONDITIONAL DISTANCE.

THEOREME-1 $\mathcal{A}$ is an element of $\mathcal{E}_0$, $\mathcal{C}$ is a σ-algebra and $(\mathcal{C}_n)_{n\in\mathbb{N}}$ increasing sequence in $\mathcal{E}_0$ so that $(\mathcal{C}_n) \nearrow \mathcal{C}$ *i.e* $\mathcal{C}= \overline{\sigma(\bigcup_n \mathcal{C}_n)}$.

Then $$H(\mathcal{A}|\mathcal{C}) = \lim_{n\to\infty} H(\mathcal{A}|\mathcal{C}_n)$$

Proof

$$H(\mathcal{A}|\mathcal{C}) \le \inf_{\substack{\mathcal{B}\subset\mathcal{C} \\ \mathcal{B}\text{ finite}}} H(\mathcal{A}|\mathcal{B}) \qquad \text{(II-4-iii)}$$

We have to prove that $H(\mathcal{A}|\mathcal{C}) = \inf_{\substack{\mathcal{B}\subset\mathcal{C} \\ \mathcal{B}\text{ finite}}} H(\mathcal{A}|\mathcal{B})$

Because $\mathcal{C} = \overline{\sigma(\bigcup_n \mathcal{C}_n)}$, we can say that:

$$\forall A\in\mathcal{C} \quad \forall\varepsilon>0 \;\; \exists n \in \mathbb{N} \quad \exists B \in \mathcal{C}_n; \quad P(A\Delta B) = \| 1_A - 1_B \|_2^2 \le \varepsilon$$

and $\bigcup_n \mathcal{L}^2(\mathcal{C}_n)$ is dense in $\mathcal{L}^2(\mathcal{C})$. The orthogonal projection of 1_A on $\mathcal{L}^2(\mathcal{C})$ is $P(A|\mathcal{C})$.

It results from the preceding that,

$$P(A|\mathcal{C}_n) \xrightarrow[n\to\infty]{} P(A|\mathcal{C}) \quad \text{and } H(\mathcal{A}|\mathcal{C}) = \lim_{n\to\infty} H(\mathcal{A}|\mathcal{C}_n)$$

THEOREME 2-

$\mathcal{A}$ and $\mathcal{B}$ being elements of $\mathcal{E}_0$, $\mathcal{C}$ a σ-algebra,

$$H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) = H(\mathcal{A}|\mathcal{B}\vee\mathcal{C}) + H(\mathcal{B}|\mathcal{C})$$

This result is immediat for $\mathcal{C} \in \mathcal{E}_0$. The general case is an immediat consequence of the Theoreme-1.

COROLLARY: For fixed $\mathcal{C} \in \mathcal{E}_0$

$$\delta(\mathcal{A},\mathcal{B}|\mathcal{C}) = \delta_{\mathcal{C}}(\mathcal{A},\mathcal{B}) = H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) - I(\mathcal{A},\mathcal{B}|\mathcal{C})$$

define a semi-distance on $\mathcal{E}_0$.

Properties of $\delta_{\mathcal{C}}$:

i) $\delta_{\mathcal{C}}(\mathcal{A},\mathcal{B}) = H(\mathcal{A}|\mathcal{B}\vee\mathcal{C}) + H(\mathcal{B}|\mathcal{A}\vee\mathcal{C})$

ii) If $\mathcal{B}\subset\mathcal{C}$ $\delta_{\mathcal{C}}(\mathcal{A},\mathcal{B}) = H(\mathcal{A}|\mathcal{C})$

iii) If $\mathcal{C}$ is the trivial σ-algebra,
$\delta_{\mathcal{C}}(\mathcal{A},\mathcal{B}) = d(\mathcal{A},\mathcal{B}) = H(\mathcal{A}|\mathcal{B}) + H(\mathcal{B}|\mathcal{A})$ is a distance on $\mathcal{E}_0$

iv) $\mathcal{A}$, $\mathcal{A}'$, $\mathcal{B}$, and $\mathcal{B}'$ being elements of $\mathcal{E}_0$, $\mathcal{C}$ an any σ-algebra,

-) $|H(\mathcal{A}|\mathcal{C}) - H(\mathcal{B}|\mathcal{C})| \le \delta_{\mathcal{C}}(\mathcal{A},\mathcal{B})$
-) $2I(\mathcal{A},\mathcal{B}|\mathcal{C}) + \delta_{\mathcal{C}}(\mathcal{A},\mathcal{B}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{B}|\mathcal{C})$
-) $|I(\mathcal{A},\mathcal{B}|\mathcal{C}) - I(\mathcal{A}',\mathcal{B}'|\mathcal{C})| \le 2[\delta_{\mathcal{C}}(\mathcal{A},\mathcal{A}') + \delta_{\mathcal{C}}(\mathcal{B},\mathcal{B}')]$

v) $\delta_{\mathcal{C}}(\mathcal{A},\mathcal{B})= 0$ define an equivalence relation on $\mathcal{E}_0$ and $\delta_{\mathcal{C}}(\mathcal{A},\mathcal{B}) = 0 \iff \mathcal{A}\vee\mathcal{C} = \mathcal{B}\vee\mathcal{C}$.

III GENERALIZATION.

We have in the precedings defined notions of entropy and mutual information for finite-type σ-algebras. We try in this section to extend these notions, by a natural way, to any sub-σ-algebra of $\mathcal{F}$, supposed to be from now a separable-type σ-algebra on Ω.

III -1) DEFINITIONS :

$\mathcal{A}$ and $\mathcal{B}$ being complete sub-σ-algebras of $\mathcal{F}$, we call :

i) mutual information between A and B given $\mathcal{C}$, the quantity

$$I(\mathcal{A},\mathcal{B}|\mathcal{C}) = \sup_{\substack{V\in\mathcal{E}_0, W\in\mathcal{E}_0 \\ V\subset\mathcal{A}\ W\subset\mathcal{B}}} I(V, W|\mathcal{C})$$

ii) entropy of $\mathcal{A}$ given $\mathcal{C}$, the quantity:

$$H(\mathcal{A}|\mathcal{C}) = I(\mathcal{A},\mathcal{A}|\mathcal{C}).$$

III -2) THEOREM:

Let $\mathcal{A} = \sigma'(\bigcup_n \mathcal{A}_n)$ where $(\mathcal{A}_n)_n$ is an increasing sequence in $\mathcal{E}_0$ and let $\mathcal{B} \in \mathcal{E}_0$ such that $\mathcal{B} \subset \mathcal{A}$. Then,

$\forall\varepsilon > 0$, $\exists\, n_0\in \mathbb{N}$ such that $H(\mathcal{B}) - \varepsilon \le H(\mathcal{A}_{n_0})$.

PROOF :

Put $\mathcal{B} = \sigma'(B_1, \ldots, B_p)$ where $(B_i)_{i=1,\ldots p}$ is a partition of Ω. It is known ([7]) that:

$\forall \delta > 0, \forall i = 1,\ldots, p, \exists A'_i \in \bigcup_n \mathcal{A}_n$ such that $P(B_i \Delta A'_i) < \frac{\delta}{2^i}$.

Put $A_1 = A'_1$, $A_i = A'_i \setminus \bigcup_{j=1}^{i-1} A_j$ $\forall i = 2, \ldots, p$ and $A_{p+1} = (\bigcup_{i=1}^{p} A_i)^C$

$(A_i)_{i\in[1\ldots p+1]}$ is a partition of Ω which elements are contained in one of the $\mathcal{A}_n$, say $\mathcal{A}_{n_0}$, and we have:

$$\forall i,\ 1\le i \le p,\quad P(B_i \Delta A_i) < \delta,\ P(A_{p+1}) < \delta$$

Morever, $\forall i,\ 1 \le i \le p,\ |P(Bi) - P(Ai)| \le P(B_i \Delta A_i) < \delta$.

Using continuity of $H(\mathcal{B})$ with respect to $P(B_1), \ldots, P(B_p)$, it is easy to show:

$\forall \varepsilon > 0, \exists \eta > 0, \forall i,\ 1 \le i \le p,\ |P(B_i)-P(A_i)| < \eta$ and $P(A_{p+1})<\delta$

Then, $|H(\mathcal{B}) - H(\sigma(A_1, \ldots, A_{p+1}))| < \varepsilon$

Choose $\delta = \eta$, the theorem follows from,

$$H(\sigma(A_1, \ldots, A_{p+1})) \le H(\mathcal{A}_{n_0})$$

and the preceding inequality.

From the above theorem we deduce the following corollary, very useful for calculations of entropy and information for every complete sub-σ-algebras contained in $\mathcal{F}$ (see I-2).

COROLLARY 1:

If $\mathcal{A} = \sigma'(\bigcup_n \mathcal{A}_n)$ and $\mathcal{B} = \sigma'(\bigcup_n B_n)$.

with $(\mathcal{A}_n)_n$ (resp. $(\mathcal{B}_n)_n$) increasing sequences of finite σ-algebras, then

i) $H(\mathcal{A}) = \lim_n\uparrow H(\mathcal{A}_n)$

ii) $I(\mathcal{A},\mathcal{B}) = \lim_{n,m}\uparrow\uparrow I(\mathcal{A}_n, \mathcal{B}_m) = \lim_n\uparrow I(\mathcal{A}_n, \mathcal{B}_n)$

All definitions and properties in section II cand be extended without difficulties and without modification to the set:
$\mathcal{E}_1 = \{ \mathcal{A} \in \mathcal{E},\ H(\mathcal{A}) < +\infty \}$, finite sums being subtituted with absolutely convergent series.

III-3) PROPOSITION :

Let $\mathcal{A}$ be any complete sub-σ-algebra of $\mathcal{F}$, then if $H(\mathcal{A})$ is finite, $\mathcal{A} \in \mathcal{E}_1$.

PROOF:

Put $\mathcal{A} = \sigma'(\bigcup_n \mathcal{A}_n)$ with $(\mathcal{A}_n)_n$ an increasing sequence of finite σ-algebras.

Denote by :

*) $(A_n^i)_{i\in I_n}$ a partition generating $\mathcal{A}_n$;

*) $\zeta_{\mathcal{A}_n} = \sum_{i\in I_n} (\log \frac{1}{P(A_n^i)}).1_{A_n^i}$

Notice that $E(\zeta_{\mathcal{A}_n}) = H(\mathcal{A}_n)$, $(\zeta_{\mathcal{A}_n})$ is an increasing sequence and

$\forall x \in \Omega \quad \zeta_{\mathcal{A}_n}(x) = \log \frac{1}{P(A_n^{i(x)})}$ where $A_n^{i(x)}$ is the element of the partition $(A^i)_{i\in I_n}$ wich contains x.

From I-3), $\forall x \in \Omega$:

*) if x belongs to an atom, $\lim_{n\to\infty} \zeta_{\mathcal{A}_n}(x) = \zeta(x) < +\infty$

*) if x does not belong to an atom, $\zeta(x) = +\infty$.

Then, if $\mathcal{A} \notin \mathcal{E}$, the complementary set of the set of atoms is non-null and, $\lim_n H(\mathcal{A}_n) = H(\mathcal{A}) = E(\zeta) = +\infty$.

III-4) THEOREM :

For each $\mathcal{A}$, $\mathcal{B} \in \mathcal{E}_1$, put :

$d(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}|\mathcal{B}) + H(\mathcal{B}|\mathcal{A})$, then we define a distance on $\mathcal{E}_1$, and the space $(\mathcal{E}_1, d)$ is a complete metric space.

Morever $(\mathcal{E}_0, d)$ is dense in $(\mathcal{E}_1, d)$.

PROOF:

d is a distance on $\mathcal{E}_1$ as on $\mathcal{E}_0$. Consider now a Cauchy sequence $(\mathcal{A}_n)_n$ in $(\mathcal{E}_1, d)$ such that

$$d(\mathcal{A}_{n+1}, \mathcal{A}_n) < \frac{1}{2^n}.$$

$$d(\bigvee_{k=0}^{p+1} \mathcal{A}_{n+k}, \bigvee_{h=0}^{p} \mathcal{A}_{n+h}) = H(\bigvee_{k=0}^{p+1} \mathcal{A}_{n+k} | \bigvee_{h=0}^{p} \mathcal{A}_{n+h}) =$$

$$H(\mathcal{A}_{n+p+1} | \bigvee_{h=0}^{p} \mathcal{A}_{n+h}) \le H(\mathcal{A}_{n+p+1}|\mathcal{A}_{n+p}) \le d(\mathcal{A}_{n+p+1}, \mathcal{A}_{n+p}).$$

But we have the following :

$$d\left(\bigvee_{h=0}^{p} \mathcal{A}_{n+h}, \mathcal{A}_n\right) \leq \sum_{k=0}^{p-1} d(\mathcal{A}_{n+k+1}, \mathcal{A}_{n+k}) \leq \sum_{k=0}^{p-1} \frac{1}{2^{n+k}} < \frac{1}{2^{n-1}}$$

Then, $H(\bigvee_{k=0}^{p} \mathcal{A}_{n+k}) - H(\mathcal{A}_n) < \frac{1}{2^{n-1}}$.

We deduce that $(\bigvee_{k=0}^{n} \mathcal{A}_k)_n$ is an increasing sequence of sub-σ-algebras which converge to $(\bigvee_{k=0}^{\infty} \mathcal{A}_k)$,

that $\mathcal{B}_n = (\bigvee_{k=n}^{\infty} \mathcal{A}_k) \in \mathcal{E}_1$.

and that the two sequences $(\mathcal{B}_n)_n$ and $(\mathcal{A}_n)_n$ have the same limit.

As $(\mathcal{B}_n)$ is a decreasing sequence in $\mathcal{E}_1$, it has a limit in $\mathcal{E}_1$

We show now that this limit is :

$$\mathcal{B} = \bigcap_n \mathcal{B}_n = \lim_n \sup \mathcal{A}_n = \bigcap_n (\bigvee_{k \geq n} \mathcal{A}_n)$$

Let $(\mathcal{B}_k^n)_{k \in I_n}$ be a partition generating $\mathcal{B}_n$ wich is finer than $(\mathcal{B}_{n+1}^k)_{k \in I_{n+1}}$ (this is possible because $(\mathcal{B}_n)_n$ is decreasing).

$\forall x \in \Omega$, the atom of $\mathcal{B}$ which contains x is nothing else that $\bigcup_n B_n^{k(x)}$, where $B_n^{k(x)}$ is the element of the partition $(\mathcal{B}_n)$ which contains x. $(B_n^{k(x)})_n$ is an increasing sequence and then the sequence $(\zeta_{\mathcal{B}_n})_n$ converges to $\zeta_{\mathcal{B}}$.

$$H(\mathcal{B}_n) = E(\zeta_{\mathcal{B}_n})_{\overrightarrow{n \to \infty}} E(\zeta_{\mathcal{B}}) = H(\mathcal{B}).$$

As for each n $\mathcal{B}_n \subset \mathcal{B}$, $d(\mathcal{B}_n, \mathcal{B}) = H(\mathcal{B}_n) - H(\mathcal{B})$ and

$$d(\mathcal{B}_n, \mathcal{B}) \xrightarrow[n \to +\infty]{} 0$$

The density of $\mathcal{E}_0$ in $\mathcal{E}_1$ is easy.

Remarks: i) if $\mathcal{A}, \mathcal{B} \in \mathcal{E}_1$ then $\mathcal{A} \vee \mathcal{B} \in \mathcal{E}_1$

ii) if $\mathcal{C}$ and $\mathcal{C}'$ are any σ-algebras and if $\mathcal{C} \subset \mathcal{C}'$ then $\mathcal{E}_1(\mathcal{C}) \subset \mathcal{E}_1(\mathcal{C}')$

III-5 REMARKS

1) Differential Entropy.

P is a probability distribution with a step-density function f with respect to Lebesque measure μ; f associated with the values α_i $i \in \mathbb{N}$.

The differential entropy is defined by

$$H(P) = \int f \mathrm{Log}(f) d\mu$$

and if we note

$$H(\sigma(f)) = - \sum_i P(f=\alpha_i) \mathrm{Log}\ P(f=\alpha_i)$$

we have the relation

$$H(P) + H(\sigma(f)) = \sum_i P(f=\alpha_i) \mathrm{Log} \frac{1}{\mu(f=\alpha_i)} .$$

Examples can be built for wich the differential entropy is infinite (resp. finite) and $H(\sigma(f))$ is finite (resp. infinite).

Therefore those two notions are different; none of them are the generalization of the other.

2) Entropy of a sequence of random variables.

The classical definition of the entropy of a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ is

$$\lim_n \frac{1}{n} H(X_1, X_2, \ldots, X_n)$$

where

$$H(X_1, X_2, \ldots, X_n) = H(\bigvee_{i=1}^{n} \sigma(X_i))$$

A necessary condition for this limite to be finite is that

$\forall i,\ \sigma(X_i) \in \mathcal{E}_1$, i.e consequently: $\forall n$ $H(X_1, X_2, \ldots, X_n)$ is finite.

We proved above that $H(\bigvee_{i=1}^{\infty} \sigma(X_i)) = \lim_{n \to \infty} H(\bigvee_{i=1}^{n} \sigma(X_i))$.

Obviously $\lim_n \frac{1}{n} H(X_1, X_2, \ldots, X_n)$ is a notion which does not depend on the σ-algebra $\bigvee_{i=1}^{\infty} \sigma(X_i)$, but depend only on the sequence $X_1, X_2, \ldots, X_n$.

III-6 THEOREME:

Given $\mathcal{A}$ and $\mathcal{C}$ any two σ-algebras, $(A_i)_{i\in\mathbb{N}}$ (resp. $(C_k)_{k\in\mathbb{N}}$) the set of the atoms of $\mathcal{A}$ (resp. $\mathcal{C}$) if:

$$P\left[(\bigcup A_i)^C \cap (\bigcup_k C_k)\right] \neq 0, \text{ then } H(\mathcal{A}|\mathcal{C}) = +\infty$$

We define $H(\mathcal{A}|\mathcal{C})$ when $\mathcal{A}$ is a finite σ-algebra, the generalization of this notion is easy for an any σ-algebra $\mathcal{A}$ by means of limit theorems.

If $\mathcal{C} = \mathcal{F}_0$ we find the result of the proposition III-3:

$$\text{if} \quad P\left[(\bigcup A_i)^C\right] = 0, \text{ then } H(\mathcal{A}|\mathcal{C}) = +\infty$$

Proof: In a first time let us prove that when $\mathcal{A}$ is a finite σ-algebra generated by $(A_i)_i$:

$$H(\mathcal{A}|\mathcal{C}) \geq \sum_k H(\mathcal{A}|C_k).P(C_k) \quad \text{where } H(\mathcal{A}|C_k) = \sum_i \varsigma\left(P(A_i|C_k)\right)$$

this result can be extented to the case where $\mathcal{A}$ is an any σ-algebra. These extension is obtained by means of limit theorems.

We have:

$$P(A_i|\mathcal{C}).\mathrm{Log}[P(A_i|\mathcal{C})]\,\big|_{C_k} = P(A_i|C_k).\mathrm{Log}[P(A_i|C_k)]$$

so that

$$H(\mathcal{A}|\mathcal{C}) \geq \sum_k H(\mathcal{A}|C_k)\ 1_{C_k}$$

and

$$H(\mathcal{A}|\mathcal{C}) \geq \sum_k H(\mathcal{A}|C_k).P(C_k)$$

Let us suppose now that:

$$P\left[(\bigcup A_i)^C \cap (\bigcup_k C_k)\right] \neq 0, \text{ then}$$

$$\exists\, k_0;\quad P\left[(\bigcup A_i)^C \cap C_{k_0})\right] \neq 0$$

The same demonstration for the proposition III-3 but with the probability $P_{C_{k_0}}$ conditioned by C_{k_0} prove that $H(\mathcal{A}|C_{k_0}) = +\infty$

We deduce that $H(\mathcal{A}|\mathcal{C}) = +\infty$

CONCLUSION

Mutual information between any two random variables can be defined as mutual information between their induced σ-algebras. Then we have a new tool for statistical study of a qualitative random variables.

Practical applications of this point of vew are numerous, in particular in biology.

BIBLIOGRAPHIE

[1] H. CAUSSINUS *Projections révélatrices- Modèles pour l'Analyse des onnées multidimensionnelle* - A.S.U Economica

[2] G. CHOQUET: Théory of *capacities* Anales Institut de Fourier Grenoble 5, pages 131-295 (1953-1954).

[3] C. DELACHERIE, P.A MEYER: *Probabilités et Potentiel*, Tome 1 Hermina (1980)

[4] T. DUCAN: *On the calculation of the mutual information*, SIAM Journal of Math 19 pages 215-220 (1990).

[5] T.T. KADOTA, M. ZAKAI, J. ZIV: *Mutual information of the white Gaussian chanel with and without feedback*, IEEE Trans. Inf. Theory Vol I t-17, pages 368-371 (1971).

[6] A.I. KHINCHIN: *Mathematical foundations of information theory*. Dover publications, New York (1957).

[7] M. LOEVE: *Probability theory* 4th Edition Springer-Verlag N.Y- Heidelberg- Berlin.

[8] P.A. MEYER: *Probabilité et Potentiel*

[9] J. NEVEU: *Bases mathématiques du calcul des probabilités* Masson (1970).

[10] M. Pinsker: *Information and information stability of random variables and process*. Translated from Russian and edited by A. FEISTEIN, San Francisco, Holder Day (1964).

[11] Ya. G. SINAI: *Introduction to ergodic theory*. Mathematical notes. Princeton University Press (1976).

[12] E. ZEIDLER *Nonlinear functional aalysis and its applications*. Tom III *Variational méthods and optimization*. Springer-Verlag (1985).