

50376
1974
63

N° d'ordre : 304

50376
1974
63

THESES

présentées à

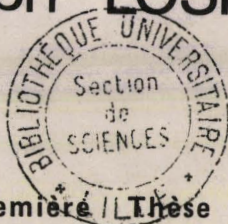
L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

pour obtenir

le grade de Docteur Es Sciences Mathématiques

par

Joseph LOSFELD.



Première Thèse :

Information fournie par un ensemble d'observateurs
et applications aux questionnaires
et à l'analyse des données

Deuxième Thèse :

Equivalence d'Expressions Rationnelles

Thèses soutenues le 17 Juin 1974 devant la Commission d'Examen

MM. J. KAMPE DE FERIET, Président d'honneur - Rapporteur
P. POUZET, Président
D. BOSQ, Examinateur
M. NIVAT, Invité
C. PICARD, Invité

A ma femme,

à François, Caroline et Anne-Sophie.

Je tiens à exprimer ma respectueuse reconnaissance à Monsieur le Professeur KAMPE DE FERIET, président d'honneur de cette thèse et directeur de ce travail, pour les précieux conseils qu'il n'a cessé de me prodiguer. Ce mémoire n'aurait jamais vu le jour sans l'aide qu'il m'a apportée ni le temps qu'il a bien voulu m'accorder.

Que Monsieur le Professeur POUZET, qui a accepté la présidence du Jury, trouve ici l'expression de ma gratitude pour ses conseils et pour l'intérêt qu'il manifeste depuis de nombreuses années à l'évolution de mes travaux d'enseignement et de recherche.

Je tiens à exprimer ma profonde reconnaissance à Monsieur le Professeur NIVAT qui m'a proposé un deuxième sujet auquel j'ai porté un vif intérêt.

Je remercie également Monsieur le Professeur PICARD et Monsieur le Professeur BOSQ de m'avoir fait l'honneur de participer au Jury.

Je remercie mon ami Gérard COMYN pour sa précieuse collaboration. Je souhaite que ce travail ne soit qu'une première étape dans nos activités de recherche communes.

Je tiens également à remercier le Secrétariat de l'U.E.R. de Mathématiques de l'Université des Sciences et Techniques de LILLE et notamment Mesdames BERAT et LENGAGNE, ainsi que Madame DESCARPENTRIES, du Secrétariat du Département Informatique de l'I.U.T., qui ont permis la réalisation matérielle de cet ouvrage.

INTRODUCTION

PREMIERE PARTIE

INFORMATION SUR UN EVENEMENT

- | | |
|--------------|--|
| Chapitre I | INFORMATION APPOREE PAR UN GROUPE D'OBSERVATEURS
SUR UN EVENEMENT |
| Chapitre II | INFORMATION SUR UN ESPACE PRODUIT |
| Chapitre III | DEPENDANCE ET INDEPENDANCE DES OBSERVATEURS |
| Chapitre IV | CONSTRUCTION D'UNE INFORMATION APPOREE PAR UN COUPLE
D'OBSERVATEURS |
| Chapitre V | CONCLUSIONS DE LA PREMIERE PARTIE |

DEUXIEME PARTIE

INFORMATION SUR UNE PARTITION

- | | |
|--------------|--------------------------|
| Chapitre VI | INFORMATION MOYENNE |
| Chapitre VII | EXEMPLES ET APPLICATIONS |

CONCLUSION

BIBLIOGRAPHIE

INTRODUCTION

Il n'est pas dans notre intention d'analyser les différents sens que l'on donne au terme "information" dans la vie courante, ni les différents modèles mathématiques qui en ont été proposés ; des auteurs plus qualifiés ont étudié ces problèmes avec pertinence [5], [19], [32], [38], [39]. Nous voulons simplement, dans cette introduction, montrer, sur des exemples concrets, pourquoi les formalisations théoriques habituelles ne permettent pas, dans certains cas, de quantifier de façon satisfaisante diverses notions intuitives d'information, et présenter les solutions que nous proposons dans ce mémoire.

Si l'on caractérise une personne par le fait que :

"C'est un HOMME"

"C'est un HOMME MARIE"

....

"C'est un HOMME MARIE, de TRENTE ANS, né à LILLE,
père de TROIS ENFANTS"

on augmente manifestement l'"information" que l'on possède sur cet individu au fur et à mesure que l'on *accumule les renseignements*.

Pour rendre compte de ce phénomène dans le cadre de la théorie classique de l'information de WIENER et SHANNON, ou dans celui de la théorie généralisée de l'information sur un événement de J. KAMPE DE FERIET et B. FORTE, on peut, par exemple, définir dans l'ensemble Ω des Français, les sous-ensembles :

A = sous-ensemble des HOMMES

B = sous-ensemble des HOMMES MARIÉS

C = sous-ensemble des HOMMES MARIÉS, de TRENTE ANS

et remarquer que, par construction, l'ensemble C est inclus dans l'ensemble B, lui-même inclus dans l'ensemble A,

On associe alors à l'ensemble A (resp. B, C) le nombre réel non négatif $I(A)$ (resp. $I(B), I(C)$) qui est supposé quantifier l'information apportée par la réalisation de l'événement : "*cette personne appartient à l'ensemble A (resp. B, C)*". Ces quantités vérifient :

$$I(C) \geq I(B) \geq I(A).$$

En effet, *cette propriété de l'information classique :*

$$B \subseteq C \implies I(B) \geq I(C)$$

est l'axiome essentiel de la théorie de l'information généralisée.

Nous distinguerons trois étapes dans le raisonnement précédent :

- Intuitivement, pour caractériser un individu ω_0 de Ω , la proposition $P_B = \text{"C'est un HOMME MARIE"}$ est *plus précise*, donc apporte plus d'information que la proposition $P_A = \text{"C'est un HOMME"}$; nous noterons $P_B \rightarrow P_A$ (P_B implique P_A).
- Une *conséquence* de ce fait est que l'ensemble B des HOMMES MARIÉS est *inclus* dans l'ensemble A des HOMMES : $B \subseteq A$.
- Les modèles classiques sont construits sur cette seconde remarque :
comme B est inclus dans A , l'affirmation " ω_0 appartient à B " apporte plus d'information sur la *localisation* de l'individu ω_0 dans Ω que l'affirmation " ω_0 appartient à A ". Dans ce cadre, l'information est maximum quand on a réussi à identifier l'élément ω_0 , c'est-à-dire quand on peut affirmer qu'il appartient à un ensemble formé d'un seul élément : $\omega_0 \in \{\omega_0\}$.

Résumons cette démarche dans le schéma suivant :

Substitution de l'inclusion
des ensembles associés à
l'implication des propo-
sitions logiques

$$\left\{ \begin{array}{c} P_B \rightarrow P_A \\ \Downarrow \\ B \subseteq A \\ \Downarrow \\ I(B) \geq I(A) \end{array} \right.$$

Axiome de base de la
théorie généralisée de
l'information sur les
événements.

Montrons, sur deux exemples simples, que ce schéma n'est pas complètement satisfaisant quand on donne aux ensembles A et B une signification concrète.

- 1) Si, parlant d'un français, j'affirme que c'est un PROFESSEUR, par ailleurs CLOWN PROFESSIONNEL, j'apporte sur cette personne une information $I(A)$, où A représente l'ensemble des PROFESSEURS, CLOWNS PROFESSIONNELS. Si je dis qu'il appartient à l'ensemble B des PROFESSEURS, CLOWNS PROFESSIONNELS, HOMMES MARIÉS, de TRENTE ANS, j'apporte une information $I(B)$ manifestement supérieure à $I(A)$. Dans la réalité de 1974, il se trouve que l'ensemble A est réduit à un seul français : ω_0 , parfaitement identifiable. De l'égalité $A = B = \{\omega_0\}$ on doit déduire : $I(A) = I(B)$, ce qui traduit effectivement le fait que l'appartenance à l'ensemble B ne permet pas de *localiser* avec plus de précision cette personne dans l'ensemble des français que l'appartenance à l'ensemble A. Pourtant, il est clair que l'affirmation correspondant à l'ensemble B *caractérise* avec plus de précision cette personne que l'affirmation correspondant à l'ensemble A.

Pour que cette *meilleure caractérisation* s'accompagne d'une *meilleure localisation*, il est nécessaire de se placer dans le cadre d'un *ensemble hypothétique de français "possibles"* dans lequel A et B ne sont pas réduits à un seul élément. C'est d'ailleurs parce qu'il est potentiellement possible d'avoir un PROFESSEUR CLOWN PROFESSIONNEL qui soit FEMME, CELIBATAIRE, de VINGT CINQ ANS que la seconde affirmation apporte plus d'information que la première.

- 2) Soit une souche bactérienne parfaitement identifiée, répertoriée, par exemple une souche de référence internationale : PSEUDOMONAS FLUORESCENS, C.I.P. (Collection de l'Institut Pasteur), N° 5690, qui est conservée sous vide, en ampoule scellée, à l'état lyophilisé. Supposons qu'un laboratoire mette au point un nouveau test constituant un critère biochimique. Avant toute expérience, il est impossible de connaître la réaction de la souche à ce test ; il est clair que la réalisation matérielle de l'expérience (par exemple, la souche C.I.P. N° 5690 réagit positivement au test), apportera une information qui permettra de *caractériser* avec plus de précision les propriétés de l'individu étudié. On ne peut lier cette information à une *meilleure localisation* car la souche bactérienne est déjà identifiée et que l'on cherche seulement à en décrire les propriétés.

Ces deux exemples nous montrent que si l'on attache à l'ensemble Ω une signification concrète, une meilleure caractérisation d'un élément $w \in \Omega$ peut ne pas être associée à une meilleure localisation de cet élément dans la population réellement étudiée. C'est-à-dire, comme nous l'avons vu, que si l'implication stricte : $(P_B \rightarrow P_A)$ équivaut à ($\hat{B}$ strictement inclus dans $\hat{A}$) dans un certain espace abstrait, il est possible dans les faits que les réalisations B et A de $\hat{B}$ et $\hat{A}$ dans la population Ω étudiée soient égales.

La solution la plus générale de ce problème passe par la définition et l'étude détaillée de la notion d'information dans un système de propositions logiques ; nous avons cité des travaux très intéressants qui existent dans ce domaine. Cependant, pour résoudre le type de difficultés plus particulières soulevées par ces exemples, il nous a paru intéressant de définir l'information que l'on possède sur un événement comme une fonction croissante d'une part, de la précision de la localisation de cet événement dans la population étudiée, d'autre part, du nombre et de la qualité des renseignements que l'on possède sur cet événement. Dans ce but, nous définissons, dans le premier chapitre, la notion d'information apportée par un groupe d'observateurs sur un même événement.

Nous prenons comme point de départ un ensemble Ω dont un élément w représente un événement élémentaire ; nous désignons par S la classe de parties de Ω définissant les événements observables ; nous identifions en fait un événement (observable) considéré du seul point de vue de sa réalisation, ou de sa non réalisation, au sous-ensemble des événements élémentaires qui le réalisent.

L'idée de faire dépendre la mesure de l'information apportée sur la réalisation d'un événement des observateurs de cet événement est apparue très rapidement dans le développement de la théorie généralisée de l'information. Le terme d'observateur peut être pris au sens propre : un signal est émis, chaque observateur perçoit en fait un signal plus ou moins déformé ; J. KAMPE DE FERIET écrit, par exemple : "Nous supposons que les événements observables sont perçus par des observateurs ξ , dont nous désignons l'ensemble par O ; entre la réalisation d'un événement et sa perception par un observateur s'introduit une distorsion, qui peut d'ailleurs varier d'un observateur à un autre : quand l'événement w se produit, l'observateur ξ perçoit un événement w_ξ :

$$\omega_{\xi} = T_{\xi}(\omega)$$

T_{ξ} désignant une application de l'espace des événements Ω sur l'espace Ω_{ξ} des événements observés par l'observateur ξ .

Nous donnons au terme d'observateur un sens métaphorique : O sera aussi l'ensemble des paramètres d'un système Ω ; par exemple, si $\omega \in \Omega$ définit l'état de l'atmosphère à un instant t donné, nous considérons le baromètre, le thermomètre, ..., comme autant d'observateurs de ω qui nous définissent respectivement la pression ω_1 , la température ω_2 , ..., sans pour cela que la description fournie par ces observations épuise obligatoirement toute la réalité définie par ω . De même, dans les deux exemples introduits précédemment, on retrouve les composantes classiques des modèles statistiques. On peut considérer l'individu ω comme un élément d'un ensemble Ω ; une bactérie comme un élément d'un groupe de bactéries ; ou caractériser cet individu par ses propriétés : la bactérie par son profil de réaction aux tests biologiques retenu. L'ensemble O est alors l'ensemble des caractères suivant lesquels nous étudions la population.

On peut même aller plus loin : nous traitons, dans le chapitre V, un exemple où le temps tient le rôle de l'ensemble des observateurs dans l'étude d'un système évolutif.

La formalisation de certains de ces exemples nous amène naturellement à nous interroger sur la structure de l'espace des observations à utiliser. Comme J. KAMPE DE FERIET, nous supposons que, dans certains cas, les observateurs peuvent être supposés *indépendants en structure* ; c'est-à-dire que : "si lorsque toutes les observations $\omega_{\xi} = T_{\xi}(\omega)$, sauf celle d'un observateur quelconque ξ_1 , sont déjà connues, l'observation ω_{ξ_1} peut être un point arbitraire dans Ω_{ξ_1} ". Ce qui nous conduit naturellement à représenter une observation par un point :

$$\hat{\omega} = \{\omega_{\xi} | \xi \in O\}$$

de l'ensemble produit

$$\hat{\Omega} = \prod_{\xi \in \mathcal{O}} \Omega_{\xi}$$

que nous appellerons *espace des observations*. Le chapitre II est consacré à l'étude de ce problème, très important dans la pratique.

Dans le chapitre III, nous proposons différentes méthodes de constructions d'informations, à partir de considérations concernant la dépendance ou l'indépendance des observateurs. Les résultats obtenus nous permettent de calculer l'information apportée sur un événement par un ensemble d'observateurs quand on connaît l'information apportée sur cet événement par chacun des observateurs et chacun des couples d'observateurs de ce même événement. Cela nous conduit, dans le chapitre IV, à étudier, à partir d'hypothèses particulières, l'information apportée sur un événement par un couple d'observateurs.

Le chapitre V conclut la première partie de ce travail, laquelle est consacrée à l'étude de l'information apportée par un groupe d'observateurs *sur un événement*. Ce chapitre peut être abordé directement par le lecteur qui cherche essentiellement à replacer notre étude dans le cadre de la théorie généralisée de l'information. Au terme de cette partie, il nous est en effet plus facile de situer notre travail par rapport à ceux qui ont été déjà réalisés dans ce domaine : notre étude repose essentiellement sur des hypothèses concernant la dépendance et l'indépendance des observateurs alors que le développement de la plupart des travaux antérieurs porte sur la notion de composabilité, il faut noter que *les informations que nous obtenons ne sont d'ailleurs pas composables* en général. Nous présentons aussi, dans ce chapitre, d'autres directions de travail possibles ; en particulier, nous traitons en détail un exemple qui nous paraît fondamental à deux points de vue : d'une part, il nous permet de mieux distinguer les notions *d'information dans un système statique* et *d'information dans un système dynamique*, de dégager les liens entre ces deux notions ; d'autre part, cette étude ne fait intervenir ni probabilité, ni fréquence.

Les applications de la théorie classique de l'information, telles qu'elles ont été développées par SHANNON et ses successeurs, ne dérivent pas directement de l'information apportée par un événement, mais plutôt de l'information moyenne apportée par les éléments d'une partition. Il en est de même pour certaines applications que nous proposons ; *la deuxième partie de ce travail*

est donc consacrée à l'étude des notions d'information moyenne sur une partition et d'information mutuelle entre deux partitions.

Notre but, dans cette deuxième partie, n'est pas de faire une étude exhaustive de la notion d'information apportée par une partition ou par une expérience (cette étude a été entreprise en particulier par B. FORTE). Nous cherchons essentiellement à présenter quelques perspectives d'applications déduites des propriétés de l'information moyenne d'une partition finie Π de Ω .

Dans le chapitre VI, nous définissons et étudions les notions d'information moyenne apportée sur une partition par un ensemble d'observateurs dans l'espace Ω des événements d'une part, et dans l'espace $\hat{\Omega}$ des observations d'autre part ; ces résultats se déduisent directement de la formule de définition et des propriétés des informations définies dans la première partie.

L'exemple numérique que nous traitons en détail dans le chapitre VII nous permet d'illustrer les résultats précédents et de montrer comment ces résultats englobent et généralisent ceux de la théorie classique de SHANNON ; en particulier, ils indiquent clairement quelles solutions nous apportons aux problèmes soulevés dans cette introduction. Dans ce dernier chapitre, et à partir de l'exemple traité, nous proposons différentes directions d'applications possibles de ces notions en analyse des données, en classification automatique, dans les questionnaires.

PREMIERE PARTIE

INFORMATION SUR UN EVENEMENT

CHAPITRE I

INFORMATION APPORTEE PAR UN GROUPE
D'OBSERVATEURS SUR UN EVENEMENT

La première partie de ce chapitre est consacrée aux rappels des différentes définitions et propriétés de la mesure généralisée de l'information dont nous avons besoin par la suite. Nous précisons en même temps les notations et les conventions d'écriture que nous avons adoptées. Dans la seconde partie de ce chapitre nous présentons en détail des définitions originales concernant la notion d'information apportée par un groupe d'observateurs sur un événement.

I.1.- RAPPELS ET NOTATIONS.

Notre démarche prolonge celles de J. KAMPE DE FERIET et B. FORTE, notamment sur la mesure généralisée de l'information sur un événement [12] [13], sur la mesure de l'information par un ensemble d'observateurs indépendants [14] [15] [17] et sur les "collecteurs" [8]. Nous utiliserons donc la terminologie et les notations développées dans ces travaux.

I.1.1.- INFORMATION SUR UN EVENEMENT.

Soient Ω un ensemble quelconque, non vide, d'événements élémentaires ω , S une σ -algèbre de Boole de parties de Ω (nous dirons aussi une tribu); l'espace mesurable (Ω, S) sera appelé *espace des événements* ou *espace des phases*.

Dans la suite, chaque fois que nous parlerons de "*mesure de l'information*" ou plus brièvement d'"*information*" *apportée par (ou sur) un événement observable (ou événement)* en utilisant la lettre I , il s'agira d'une fonction d'ensemble définie sur une classe de parties de Ω (en général la σ -algèbre S) satisfaisant aux axiomes 1,2,3 et aux définitions suivantes qui furent introduits par

J. KAMPE DE FERIET et B. FORTE [12], [13] :

AXIOME 1 : L'information I est une application de S dans $\bar{\mathbb{R}}^+$

AXIOME 2 : L'information est monotone pour l'inclusion :

$$(A, B) \in S \times S \text{ et } B \subseteq A \Rightarrow I(B) \geq I(A)$$

les valeurs universelles sont

$$I(\Omega) = 0 \text{ et } I(\emptyset) = +\infty$$

Notons pour tout $A \in S$

$$A^{(i)}, i = 0, 1 : A^{(0)} = A \text{ et } A^{(1)} = A^c$$

A^c étant le complémentaire de A dans Ω .

AXIOME 3 : Les événements A et B de S sont indépendants pour l'information I (nous dirons aussi : I -indépendants), si I satisfait aux quatre conditions :

$$I(A^{(i)} \cap B^{(j)}) = I(A^{(i)}) + I(B^{(j)}) \quad i, j = 0, 1$$

Etant donné les valeurs adoptées pour l'information apportée par Ω et $\emptyset$, l'indépendance des événements A et B est équivalente à l'indépendance des algèbres de Boole $\mathcal{A}$ et $\mathcal{B}$ engendrées respectivement par A et B . D'une manière générale nous dirons que :

I.1.1.- a) DEFINITION.

Les algèbres de Boole (ou les σ -algèbres de Boole) de la famille

$$\{A_i \mid i \in K, A_i \subseteq S\}$$

sont I -indépendantes en information (respectivement I - σ -indépendantes) si pour tout sous-ensemble fini (resp. dénombrable) $K' \subseteq K$ d'indices et pour tout choix de $A_i \in \mathcal{A}_i$, on a :

$$I\left(\bigcap_{i \in K'} A_i\right) = \sum_{i \in K'} I(A_i)$$

Comme l'a montré J. KAMPE DE FERIET dans [13], si l'on exclut les ensembles de S dont les valeurs d'information sont 0 ou $+\infty$, il est nécessaire pour que les notions d'indépendance en information aient un sens que *les algèbres de la famille* $\{A_i \mid i \in K, A_i \subseteq S\}$ *soient indépendantes en structure* (nous disons aussi *M-indépendantes* pour employer la terminologie de D.A. KAPPOS [20]) : c'est-à-dire que *pour tout sous-ensemble fini* $K' \subseteq K$ *d'indices et pour tout choix de* A_i *satisfaisant à :*

$$A_i \in \mathcal{A}_i, A_i \neq \emptyset, i \in K'$$

il faut que

$$\bigcap_{i \in K'} A_i \neq \emptyset$$

On parlera de σ -M-indépendance quand K' est dénombrable. Nous emploierons très souvent cette notion d'indépendance par la suite ; intuitivement étant donné $(A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2$, si A_1 et A_2 sont deux sous-algèbres de Boole de S indépendantes en structure, A_1 ne peut en aucun cas contenir A_2 ou A_2^c , et inversement.

I.1.1.- b) DEFINITION.

L'information conditionnelle fournie par l'événement $B \in S$, *lorsque l'on sait déjà que* $A \in S$ *est réalisé, est définie par :*

$$I(B/A) = I(A \cap B) - I(A) \text{ si } I(A) < +\infty$$

Quand $I(A) = +\infty$, le second membre de cette relation est indéterminé ; nous leverons cette indétermination en supposant que :

$$I(B/A) = I(B) \text{ si } I(A) = +\infty$$

Ce choix est le seul qui rend équivalentes les deux propositions :

"Les événements A *et* B *sont* I -*indépendants"*

et $I(B/A) = I(B)$

Avec cette convention, pour tout A fixé, on montre que l'application $I(. / A) : S \rightarrow \mathbb{R}^+$, vérifie les axiomes 1 et 2, ce qui justifie la terminologie employée : $I(. / A)$ est une véritable information sur (Ω, S) . Notons en particulier pour que tout $A \in S$

$$. \quad I(A/\emptyset) = I(A/\Omega) = I(A)$$

$$. \quad I(\emptyset/A) = +\infty \quad \text{et} \quad I(\Omega/A) = 0$$

$$. \quad I(A) = +\infty \iff \forall B \in S, \quad I(A/B) = +\infty$$

I.1.1.- c) DEFINITION.

Soit $(A, B) \in S \times S$; on appelle *information mutuelle* entre A et B la quantité :

$$I(A; B) = I(A) + I(B) - I(A \cap B)$$

$$\text{si } I(A) < +\infty \text{ et } I(B) < +\infty$$

Notons que le terme d'information mutuelle, consacré par l'usage [3], [27], peut prêter à confusion. En aucune manière l'information mutuelle de deux événements n'est une information ; en effet cette quantité n'est pas nécessairement positive

$$I(A; B) \in \mathbb{R} \cup \{-\infty\}$$

de plus elle n'a aucune propriété de monotonie, du type de celle de l'axiome 2.

Quand $I(A) = +\infty$ et/ou $I(B) = +\infty$ le second membre de cette relation est indéterminé ; nous leverons cette indétermination en supposant que

$$I(A; B) = 0 \text{ si } I(A) \text{ et/ou } I(B) = +\infty$$

Ce choix est le seul qui rend équivalentes les deux propositions
 "Les événements A et B sont I -indépendants"

$$\text{et} \quad " I(A; B) = 0 "$$

La plupart des travaux en théorie généralisée de l'information que nous avons cités, ou qui sont indiqués dans la bibliographie, portent pour l'essentiel sur l'étude de la notion de composabilité de l'information.

AXIOME 4 : Une information I est composable s'il existe une application F de $\bar{R}^+ \times \bar{R}^+$ dans $\bar{R}^+$ telle que, pour tout couple $(A,B) \in S \times S$, $A \cap B = \emptyset$, on ait :

$$I(A \cup B) = F(I(A), I(B))$$

Les informations I que nous utiliserons par la suite satisferont aux axiomes 1,2,3 et aux définitions a,b,c précédentes ; mais nous ne supposerons pas qu'elles vérifient l'axiome 4.

I.1.2.- INFORMATION SUR UN ENSEMBLE ORDONNE.

Les composantes communes de diverses définitions particulières de l'information ont été mises en relief lors des "Rencontres en Théories de l'Information" de juin 1973. D'une part J. KAMPE DE FERIET [19] a proposé d'une manière très générale de définir *l'information à partir d'un ensemble de propositions logiques* ; d'autre part nous avons indiqué [24] que l'on peut définir toute information comme une application monotone d'un ensemble T partiellement ordonné dans un ensemble totalement ordonné. Cette application, qui conserve l'ordre de tous les couples ordonnés de T , permet donc de comparer les éléments des couples non ordonnés ; la relation d'ordre sur T doit être convenablement choisie en fonction du sens intuitif que l'on cherche à associer à la notion d'information. On notera aussi sur ce sujet les travaux de J. SALLANTIN [28], [29].

Les axiomes et les définitions que nous allons proposer dans la deuxième partie de ce chapitre concernant la notion d'information apportée sur un événement par un ensemble d'observateurs et, dans le chapitre VI, d'information apportée par une partition, ne sont pas posés arbitrairement ; ils s'inspirent directement de ces bases générales que nous résumons dans l'énoncé suivant :

Soit T un ensemble partiellement ordonné pour la relation d'ordre notée $\preceq$. Une information généralisée J sur T est :

AXIOME 1 : Une application J de T dans $\bar{\mathbb{R}}^+$

AXIOME 2 : Monotone (nous la supposons décroissante)

$$(x, y) \in T \times T, x \leq y \Rightarrow J(x) \geq J(y)$$

S'il existe un majorant et un minorant universels, notés respectivement M et m , on peut éventuellement poser

$$J(M) = 0 \text{ et } J(m) = +\infty$$

AXIOME 3 : Si T est un inf-demi-treillis [4] pour la relation d'ordre, deux éléments x et y de T , d'infimum noté $x \wedge y$, sont dits indépendants s'ils vérifient :

$$J(x \wedge y) = J(x) + J(y)$$

Nous utilisons aussi les définitions suivantes :

. L'information conditionnelle fournie par l'élément $y \in T$, connaissant l'information fournie par $x \in T$, est définie par :

$$J(y/x) = J(x \wedge y) - J(x) \text{ si } J(x) < +\infty$$

$$J(y/x) = J(y) \text{ si } J(x) = +\infty$$

Cette dernière convention se justifie comme dans la définition I.1.1. - b. Notons que pour tout $x \in T$ fixé, $J(. / x)$ est une information proprement dite sur T , ce qui justifie la terminologie employée.

. L'information mutuelle entre deux éléments x et y de T est la quantité

$$J(x; y) = J(x) + J(y) - J(x \wedge y)$$

$$\text{si } J(x) < +\infty \text{ et } J(y) < +\infty$$

$$J(x; y) = 0 \text{ sinon}$$

Cette dernière convention se justifie comme dans la définition I.1.1. - c, de même il faut noter que l'information mutuelle n'est pas obligatoirement un nombre positif, qu'elle n'a aucune propriété de monotonie ; la terminologie employée, consacrée par l'usage, peut donc prêter à confusion.

I.1.3.- CONVENTION DE NOTATION.

Dans la mesure du possible nous utilisons dans ce travail les notations et le vocabulaire employés dans les travaux en théorie généralisée de l'information précédemment cités. On trouvera facilement par ailleurs (par exemple dans [25]) le détail des constructions concernant les espaces mesurables, les espaces produits etc... Nous nous contentons donc de préciser dans ce paragraphe les conventions et les notations utilisées par la suite.

I.1.3.- a) Soit Ω un ensemble quelconque, non vide, d'événements élémentaires ω , S une σ -algèbre de Boole de parties de Ω (nous dirons aussi tribu) ; l'espace mesurable (Ω, S) est appelé *espace des événements* ou *espace des phases*. Un ensemble mesurable $A \in S$ est appelé aussi *événement observable* ou plus simplement *événement*. L'espace des observateurs (O, F) est formé d'un ensemble O quelconque, non vide, d'observateurs ξ et d'une σ -algèbre de Boole F de parties de O . Un ensemble mesurable $F \in F$ sera parfois appelé un *groupe d'observateurs*. On se reportera à l'introduction de ce mémoire et aux exemples traités dans les chapitres V, VI, VII pour les diverses interprétations que l'on peut donner aux termes *observateur* et *événement*.

Nous employons la notation condensée $(\Omega \times O, S \times F)$ où :

- . $\Omega \times O$ est l'ensemble des couples (ω, ξ) pour $\omega \in \Omega$ et $\xi \in O$
- . $S \times F$ est l'ensemble des couples (A, F) pour $A \in S$ et $F \in F$

I.1.3.- b) Etant donné deux espaces mesurables quelconques (Ω_1, A_1) et (Ω_2, A_2) nous notons

- . $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$
- . $A_1 \times A_2 = \{(A, B) \mid A \in A_1, B \in A_2\}$

A chaque couple $(A, B) \in A_1 \times A_2$ nous pouvons associer le *pavé mesurable*, noté $A \times B$, qui est le sous-ensemble de $\Omega_1 \times \Omega_2$ de la forme :

$$A \times B = \{(\omega_1, \omega_2) \mid \omega_1 \in A, \omega_2 \in B\}$$

Nous notons $A_1 \otimes A_2$ la σ -algèbre de Boole de parties de $\Omega_1 \times \Omega_2$ engendrée par les pavés mesurables. L'espace mesurable produit $(\Omega_1 \times \Omega_2, A_1 \otimes A_2)$ sera parfois noté $(\hat{\Omega}, \hat{S})$.

A tout sous-ensemble $B \in A_2$ de Ω_2 , associons le *cylindre* $\hat{B}$ de base B dans Ω_2 , c'est-à-dire le sous-ensemble de $\hat{\Omega} = \Omega_1 \times \Omega_2$ de la forme $\hat{B} = \Omega_1 \times B$; l'ensemble des cylindres $\hat{B}$ de $\hat{\Omega} = \Omega_1 \times \Omega_2$, quand B parcourt A_2 , est une sous σ -algèbre Boole de $\hat{S} = A_1 \otimes A_2$, notée $\hat{A}_2$, isomorphe à la σ -algèbre A_2 de Ω_2 . Nous distinguerons de même le sous-ensemble $A \in A_1$ de Ω_1 et le cylindre $\hat{A} = A \times \Omega_2 \in \hat{A}_2$ de $\hat{\Omega} = \Omega_1 \times \Omega_2$. Les sous σ -algèbres $\hat{A}_1$ et $\hat{A}_2$ de $\hat{S}$ sont M-indépendantes (I.1.1.) car nous avons évidemment $\forall A \in A_1, A \neq \emptyset, \forall B \in A_2, B \neq \emptyset$:

$$A \times B = (A \times \Omega_2) \cap (\Omega_1 \times B) = \hat{A} \cap \hat{B} \neq \emptyset$$

Ces notations et ces résultats se généralisent facilement au cas d'une famille finie d'espaces mesurables.

I.1.3.- c) Dans le cas d'une famille quelconque d'espaces mesurables $\{(\Omega_i, A_i) \mid i \in K\}$:

Nous notons $\prod_{i \in K} \Omega_i$ le produit des ensembles, supposés non vides, $\{\Omega_i \mid i \in K\}$, c'est-à-dire l'ensemble des points $\omega = \{\omega_i \mid i \in K\}$ obtenus lorsque, pour tout $i \in K$, ω_i parcourt Ω_i .

Nous notons $\prod_{i \in K} A_i$ l'ensemble des familles $(A_i)_{i \in K} = \{A_i \mid i \in K\}$ obtenues lorsque pour tout $i \in K$, A_i parcourt A_i .

Un *pavé mesurable* est un sous-ensemble de $\prod_{i \in K} \Omega_i$ de la forme $\prod_{i \in K} A_i$ où les $A_i \in A_i$ sont supposés n'être distincts de Ω_i que pour un nombre fini de $i \in K$. Nous notons $\bigotimes_{i \in K} A_i$ la σ -algèbre de Boole de parties de $\prod_{i \in K} \Omega_i$ engendrée par les pavés mesurables.

Afin de simplifier l'écriture, quand il n'y a pas d'ambiguïté possible, nous noterons $(\prod_K \Omega_i, \otimes_K A_i)$ l'espace mesurable produit des espaces mesurables $\{(\Omega_i, A_i) \mid i \in K\}$; c'est-à-dire :

$$\prod_K \Omega_i = \prod_{i \in K} \Omega_i$$

$$\otimes_K A_i = \otimes_{i \in K} A_i$$

I.1.3.- d) Dans la suite, nous allons associer à chaque observateur $\xi \in O$ un espace mesurable formé d'un ensemble Ω_ξ non vide et d'une σ -algèbre de Boole S_ξ de parties de Ω_ξ ; les éléments de Ω_ξ (respectivement S_ξ) sont généralement notés ω_ξ (resp. C_ξ).

L'espace mesurable $(\hat{\Omega}, \hat{S})$ produit des espaces $\{(\Omega_\xi, S_\xi) \mid \xi \in O\}$ est appelé *espace des observations* :

$$\hat{\Omega} = \prod_O \Omega_\xi, \quad \hat{S} = \otimes_O S_\xi$$

les éléments de $\hat{\Omega}$ (resp. $\hat{S}$) sont généralement notés $\hat{\omega}$ (resp. $\hat{C}$).

De même, pour toute partie $F \in F$ non vide de O , nous désignons par $(\prod_F \Omega_\xi, \otimes_F S_\xi)$ l'espace mesurable produit des espaces $\{(\Omega_\xi, S_\xi) \mid \xi \in F\}$ qui est appelé *espace des observations du groupe F d'observateurs*. Les éléments de $\prod_F \Omega_\xi$ (respectivement $\otimes_F S_\xi$) sont généralement notés ω_F (resp. C_F).

Un *cylindre* de base C_F dans $\prod_F \Omega_\xi$ est un sous-ensemble de $\hat{\Omega} = \prod_O \Omega_\xi$ de la forme $C_F \times \prod_{F^c} \Omega_\xi$ que nous noterons

$$\hat{C}_F = C_F \times \prod_{F^c} \Omega_\xi$$

On sait qu'un *cylindre* $\hat{C}_F$ de base C_F dans $\prod_F \Omega_\xi$ est $\hat{S}$ -mesurable si et seulement si sa base C_F est $\otimes_F S_\xi$ -mesurable et que l'ensemble des cylindres mesurables $\hat{C}_F$ de base C_F dans $\prod_F \Omega_\xi$ est une sous σ -algèbre de $\hat{S}$ isomorphe à $\otimes_F S$. Nous noterons cette sous σ -algèbre

$$\hat{\otimes}_F S_\xi = \{ \hat{C}_F \mid \hat{C}_F = C_F \times \prod_{F^c} \Omega_\xi, C_F \in \hat{\otimes}_F S_\xi \}$$

Nous notons de même $\hat{S}_\xi$ la sous σ -algèbre de $\hat{S}$ des cylindres mesurables de base dans (Ω_ξ, S_ξ) . Remarquons en particulier que pour tout $F \in \mathcal{F}$, $F \neq \emptyset$: $\hat{\otimes}_F S_\xi = \hat{\otimes}_F \hat{S}_\xi$ ($\hat{\otimes}_F \hat{S}_\xi$ étant la sous σ -algèbre de $\hat{S}$ engendrée par l'ensemble des cylindres de $\hat{S}_\xi$, $\xi \in F$).

Soient $\hat{\omega} \in \hat{\Omega}$ et $F \in \mathcal{F}$, $F \neq \emptyset$, nous pouvons noter l'élément $\hat{\omega}$ sous la forme (ω_F, ω_{F^c}) ; étant donné $\hat{C} \in \hat{S}$ nous appelons *projection de $\hat{C}$ sur l'espace $\prod_F \Omega_\xi$* le sous-ensemble de $\prod_F \Omega_\xi$, noté C_F , défini par :

$$C_F = \{ \omega_F \in \prod_F \Omega_\xi \mid (\omega_F, \omega_{F^c}) \in \hat{C} \}$$

Dans la mesure où nous faisons l'hypothèse que pour tout couple $(\hat{C}, F) \in \hat{S} \times \mathcal{F}$, $F \neq \emptyset$, la projection de $\hat{C}$ sur $\prod_F \Omega_\xi$ est $\hat{\otimes}_F S_\xi$ -mesurable, notre notation se justifie; $\hat{C}_F$ désigne alors le cylindre de $\hat{\Omega}$ dont la base est C_F .

(Rigoureusement la projection de $\hat{C}$ sur $\prod_F \Omega_\xi$ devrait se noter différemment, par exemple $\hat{C}_F$, et le cylindre associé $\hat{\hat{C}}_F$; la convention adoptée permet de simplifier notablement l'écriture et n'entraîne aucune ambiguïté que le contexte ne puisse lever). La projection d'un $\hat{C}_i \in \hat{S}$, $i = 1, 2, \dots$ sur $\prod_F \Omega_\xi$ sera notée $(C_i)_F$ et le cylindre associé $(\hat{C}_i)_F$, $i = 1, 2, \dots$

Notons en particulier (Axiome du choix) qu'un pavé $\hat{C} = \prod_{\xi \in \mathcal{O}} C_\xi$ de $\hat{S}$ est vide si et seulement si un de ces facteurs $\{C_\xi \mid \xi \in \mathcal{O}\}$ est vide, ce qui implique pour tout $F \neq \emptyset$:

$$\hat{C} = \emptyset \iff C_F = \emptyset$$

Pour tout $(\hat{C}, F) \in \hat{S} \times \mathcal{F}$ nous avons $\hat{C}_F \supseteq \hat{C}$ et $\hat{C}_F = \hat{C}$ si et seulement si $\hat{C}$ est $\hat{\otimes}_F S_\xi$ -mesurable. Pour tout $(\hat{C}_i, F_i) \in \hat{S} \times \mathcal{F}$, $F_i \neq \emptyset$, $i = 1, 2$ nous avons :

$$\left[\begin{array}{l} \hat{C}_1 \subseteq \hat{C}_2 \\ F_1 \supseteq F_2 \end{array} \right] \Rightarrow \widehat{(C_1)}_{F_1} \subseteq \widehat{(C_2)}_{F_2}$$

en notant $(C_i)_{F_i}$, $i = 1, 2$ les projections de C_i sur $\Pi_{F_i} \Omega_\xi$, $i = 1, 2$
 et $\widehat{(C_i)}_{F_i}$, $i = 1, 2$ les cylindres associés ; l'inclusion peut être stricte
 même si :

$$\hat{C} = \hat{C}_1 = \hat{C}_2 : \hat{C}_{F_1} \subseteq \hat{C}_{F_2}$$

ou si $F = F_1 = F_2 : \widehat{(C_1)}_F \subseteq \widehat{(C_2)}_F$.

Pour tout $\hat{C}_i \times F_i \in \hat{S} \times F$, $F_i \neq \emptyset$, $i = 1, 2$, nous avons :

$$\widehat{(C_1)}_{F_1} \cap \widehat{(C_2)}_{F_2} \supseteq \widehat{(C_1 \cap C_2)}_{F_1 \cup F_2}$$

l'inclusion peut être stricte même quand $\hat{C} = \hat{C}_1 = \hat{C}_2$ ou $F_1 \supseteq F_2$ ou $F_1 \cap F_2 = \emptyset$;
 cependant si $\hat{C} = \hat{C}_1 = \hat{C}_2$ et $F_1 \supseteq F_2$:

$$\hat{C}_{F_1} = \hat{C}_{F_1} \cap \hat{C}_{F_2} = \hat{C}_{F_1 \cup F_2}$$

Notons enfin que pour tout $F_i \in F$, $F_i \neq \emptyset$, $\hat{C}_i \in \bigotimes_{F_i} S_\xi$, $i = 1, 2$, on a :

$$\begin{aligned} - \widehat{(C_1)}_{F_1} &= \hat{C}_1 \\ - \hat{C}_1 \cap \hat{C}_2 &\in \bigotimes_{F_1 \cup F_2} S_\xi \\ - \widehat{(C_1)}_{F_1} \cap \widehat{(C_2)}_{F_2} &= \widehat{(C_1 \cap C_2)}_{F_1 \cup F_2} \end{aligned}$$

Les sous-tribus de la famille $\{\hat{S}_\xi \mid \xi \in \mathcal{O}\}$ sont σ -M-indépendantes (I.1.1.). De même si $\{F_i \mid F_i \in F, F_i \neq \emptyset, i \in K\}$ est une famille quelconque d'ensembles de $\mathcal{O}$, deux à deux disjoints, les sous σ -algèbres de la famille $\{\bigotimes_{F_i} S_\xi \mid i \in K\}$ sont σ -M-indépendantes.

I.1.3.- e) Etant donné $F \in \mathcal{F}$, $F \neq \emptyset$ et $\hat{C} \in \hat{S}$, la projection C_F de $\hat{C}$ sur $\prod_F \Omega_\xi$ n'est pas nécessairement $\otimes_{F} S_\xi$ -mesurable, ce qui justifie l'hypothèse de mesurabilité de C_F que nous avons évoquée précédemment et que nous utiliserons constamment dans la suite (II.1.1.).

C. LANGRAND donne dans [21] l'exemple d'un événement mesurable d'un espace produit $(\Omega_1 \times \Omega_2, A \otimes B)$ dont la projection sur Ω_1 n'est pas A -mesurable. On connaît différentes conditions suffisantes de mesurabilités des projections [21], [25]. Remarquons enfin qu'il est trivial de montrer que si les ensembles O et $\{\Omega_\xi \mid \xi \in O\}$ sont finis, les projections sont toujours mesurables, ce qui est le cas des applications que nous proposons.

I.2.- DEFINITION DE L'INFORMATION J SUR $(\Omega \times O, S \times F)$

Soient Ω un ensemble quelconque, non vide, d'événements élémentaires ω , O un ensemble quelconque, non vide, d'observateurs ξ et S (respectivement F) une σ -algèbre de parties de Ω (resp. O).

I.2.1.- Nous cherchons, dans ce paragraphe, à définir l'information $J(A, F)$ apportée par l'ensemble des observateurs ξ du groupe $F \in \mathcal{F}$ sur la réalisation de l'événement $A \in S$ (c'est-à-dire sur la réalisation d'un événement élémentaire $\omega \in \Omega$ dont on sait seulement qu'il appartient à A). A notre sens, cette application J définie sur $S \times F$, à valeurs dans $\mathbb{R}^+$, doit posséder essentiellement les deux propriétés suivantes :

- pour un groupe $F \in \mathcal{F}$ fixé d'observateurs, $J(., F)$ est une information généralisée au sens de J. KAMPE DE FERIET et B. FORTE (I.1.1.); c'est-à-dire que la proposition " ω appartient à A " apporte une information d'autant plus grande que l'ensemble A est plus restreint.

- pour un événement $A \in S$ fixé, l'information $J(A, .)$ croît avec le nombre d'observateurs de cet événement.

Nous avons proposé (I.1.2.) de caractériser toutes les informations comme des applications dans $\mathbb{R}^+$, monotones pour une relation d'ordre partiel convenablement choisie sur l'espace de définition. Dans le cas qui nous occupe, nous travaillons évidemment sur l'ensemble des couples (A, F) de $S \times F$ sur lesquels nous définissons une relation d'ordre notée $\leq$, qui possède les propriétés exprimées ci-dessus :

$$(A_i, F_i) \in S \times F \quad i = 1, 2$$

$$\left[\begin{array}{l} A_1 \subseteq A_2 \\ F_1 \supseteq F_2 \end{array} \right] \iff (A_1, F_1) \leq (A_2, F_2)$$

Cette relation réflexive, transitive, antisymétrique, définit donc sur $S \times F$ une relation d'ordre partiel, produit des relations d'ordre $\subseteq$ sur S et $\supseteq$ sur F . L'ensemble $S \times F$ muni de cette relation d'ordre forme un treillis distributif. Notons $(. \wedge .)$ l'infimum et $(. \vee .)$ le supremum de deux éléments de $S \times F$, il vient :

$$(A_1, F_1) \wedge (A_2, F_2) = [(A_1 \cap A_2), (F_1 \cup F_2)]$$

$$(A_1, F_1) \vee (A_2, F_2) = [(A_1 \cup A_2), (F_1 \cap F_2)]$$

Le plus petit élément de $S \times F$ (au sens de $\leq$) est $(\emptyset, \emptyset)$ et le plus grand $(\Omega, \emptyset)$. Ce treillis est complété et on peut associer de manière unique à un élément $(A, F) \in S \times F$ l'élément complémentaire $(A, F)' = (A^c, F^c)$ qui vérifie les conditions exigées. Le treillis produit ainsi défini forme donc une algèbre de Boole de parties de $\Omega \times \emptyset$ différente de la σ -algèbre de Boole de parties de $\Omega \times \emptyset$ engendrée par les pavés de $S \times F$, que l'on considère généralement en probabilité.

Définissons sur le treillis $S \times F$ engendré par la relation d'ordre $\leq$ ainsi définie, une information généralisée comme nous l'avons proposé dans le paragraphe I.1.2.

I.2.2.- DEFINITION.

L'information apportée par un groupe d'observateurs $F \in \mathcal{F}$ sur un événement $A \in \mathcal{S}$, notée $J(A, F)$, est définie par une application J de $S \times F$ dans $\bar{\mathbb{R}}^+$, décroissante pour la relation d'ordre partiel $\leq$ sur $S \times F$:

$$(A_i, F_i) \in S \times F \quad i = 1, 2$$

$$(A_1, F_1) \leq (A_2, F_2) \implies J(A_1, F_1) \geq J(A_2, F_2)$$

c'est-à-dire :

$$\text{et } \left. \begin{array}{l} A_1 \subseteq A_2 \\ F_1 \supseteq F_2 \end{array} \right] \longrightarrow J(A_1, F_1) \geq J(A_2, F_2)$$

I.2.3.- DEFINITION.

Deux éléments (A_i, F_i) , $i = 1, 2$ de $S \times F$ sont indépendants en information si :

$$J[(A_1, F_1) \wedge (A_2, F_2)] = J(A_1, F_1) + J(A_2, F_2)$$

c'est-à-dire si :

$$J[(A_1 \cap A_2), (F_1 \cup F_2)] = J(A_1, F_1) + J(A_2, F_2)$$

En particulier, les événements A_1 et A_2 de S sont dits indépendants en information pour les observateurs du groupe $F \in F$ ou F -indépendants si :

$$J[(A_1 \cap A_2), F] = J(A_1, F) + J(A_2, F)$$

Les groupes d'observateurs F_1 et F_2 de F sont dits indépendants pour l'observation de l'événement $A \in S$ si :

$$J(A, (F_1 \cup F_2)) = J(A, F_1) + J(A, F_2)$$

Les seules valeurs universelles proposées par la théorie généralisée de l'information sont $J(\Omega, \emptyset) = 0$ et $J(\emptyset, \emptyset) = +\infty$. L'étude détaillée des conséquences de ces définitions va nous montrer que l'application J ainsi définie possède les propriétés exprimées en I.2.1. et nous permettre de préciser la valeur prise par J en certains éléments particuliers de $S \times F$.

I.2.4.- PROPOSITION.

Soient un événement $A \in S$ fixé et deux groupes d'observateurs F_1, F_2 de F tels que :

$$\emptyset \subseteq F_2 \subseteq F_1 \subseteq \emptyset$$

il vient :

$$0 \leq J(A, \emptyset) \leq J(A, F_2) \leq J(A, F_1) \leq J(A, \emptyset) \leq +\infty$$

Quelles valeurs donner à $J(A, \emptyset)$: information apportée sur l'événement $A \in S$ par un ensemble vide d'observateurs ?

Etant donné le sens intuitif que nous attribuons dans ce travail à $J(A, F)$, nous conviendrons de supposer que $J(A, \emptyset) = 0$ pour tout $A \in S$. Nous généraliserons d'ailleurs cette hypothèse en supposant qu'il existe une classe F_0 des groupes d'observateurs de F qui n'apportent pas d'information et pour lesquels nous poserons :

$$F_0 \in F_0, A \in S \implies J(A, F_0) = 0$$

La classe F_0 d'éléments de F contient au moins l'ensemble $\emptyset$ et peut-être réduite à ce seul élément; F_0 est héréditaire, en effet si $F_0 \in F_0$, $F \in F$ et $F \subseteq F_0$ les relations I.2.4. impliquent $F \in F_0$. Pour éviter le cas trivial où pour tout $A \in S$ et tout $F \in F$, $J(A, F) = 0$, nous supposerons qu'il existe au moins un événement $A \in S$ pour lequel $J(A, \emptyset) > 0$ c'est-à-dire que la classe $F - F_0 = \{F \in F | F \notin F_0\}$ est non vide.

Aucun problème ne se pose quant à la valeur à attribuer à $J(A, \emptyset)$. Si l'on considère d'une part que l'ensemble $\emptyset$ des observateurs peut toujours être imaginé inclus dans un ensemble plus vaste $\emptyset'$ d'observateurs qui apportent strictement plus d'information que ceux de $\emptyset$, et pour éviter d'autre part le cas limite où $J(A, \emptyset) = +\infty$ pour tout événement $A \in S$ nous conviendrons qu'il existe au moins un élément $A \in S$ tel que $J(A, \emptyset) < +\infty$.

1.2.5.- PROPOSITION.

Soient $F \in F$ un groupe fixé d'observateurs et deux événements A_1, A_2 de S tels que :

$$\Omega \supseteq A_2 \supseteq A_1 \supseteq \emptyset$$

il vient :

$$0 \leq J(\Omega, F) \leq J(A_2, F) \leq J(A_1, F) \leq J(\emptyset, F) \leq +\infty$$

L'application $J(\cdot, F)$ est donc une information généralisée au sens du paragraphe I.1.1., avec la restriction que les quantités $J(\Omega, F)$ et $J(\emptyset, F)$ ne sont pas obligatoirement égales respectivement à 0 et $+\infty$. On choisit en général ces valeurs afin qu'elles demeurent identiques pour toutes les informations et parce que ce choix rend les événements Ω et $\emptyset$ indépendants en information de tous les événements $A \in S$. Effectivement, pour tout groupe $F \in \mathcal{F}$ d'observateurs fixé, imposons que les événements Ω et $\emptyset$ soient F -indépendants en information de tous les événements de S , c'est-à-dire que, pour tout $A \in S$:

$$J((A \cap \Omega), F) = J(A, F) + J(\Omega, F)$$

$$J((A \cap \emptyset), F) = J(A, F) + J(\emptyset, F)$$

Le rapprochement de ces égalités avec les inégalités de I.2.5. implique que les seules solutions possibles sont, pour un $F \in \mathcal{F}$ donné :

- . pour tout $A \in S : J(A, F) = 0$ alors $J(\Omega, F) = J(\emptyset, F) = 0$
- . pour tout $A \in S : J(A, F) = +\infty$ alors $J(\Omega, F) = J(\emptyset, F) = +\infty$
- . sinon $J(\Omega, F) = 0$ et $J(\emptyset, F) = +\infty$

Le premier cas se rapporte aux éléments F de la classe $\mathcal{F}_0$ des groupes d'observateurs qui n'apportent pas d'information. Nous avons convenu d'exclure le second cas qui implique $J(A, \emptyset) = +\infty$ pour tout $A \in S$. Le dernier cas, qui correspond à la définition habituelle de l'information généralisée, se rapporte aux éléments F de la classe $\mathcal{F} - \mathcal{F}_0$.

Nous ne cherchons pas, par le raisonnement précédent, à démontrer la validité des valeurs universelles de l'information, mais seulement à présenter des considérations qui permettent de justifier l'énoncé suivant :

I.2.6.- DEFINITION.

La classe F_0 des groupes d'observateurs qui n'apportent pas d'information est un sous-ensemble de F tel que :

- . $\emptyset \in F_0$
- . F_0 est héréditaire : $F_0 \in F_0$, $F \in F$ et $F \subseteq F_0 \implies F \in F_0$
- . $F - F_0 = \{F \in F \mid F \notin F_0\}$ non vide.

L'information J sur $S \times F$ définie en I.2.2. prend les valeurs universelles suivantes :

$$F \notin F_0 \quad J(\Omega, F) = 0 \quad \text{et} \quad J(\emptyset, F) = +\infty$$

$$F \in F_0 \quad J(\Omega, F) = J(\emptyset, F) = 0$$

Nous définissons par la suite F_0 comme la classe des ensembles F -mesurables négligeables pour une certaine mesure finie sur $(0, F)$. Dans ce cas F_0 est stable pour l'opération de réunion dénombrable et forme donc un σ -idéal. Par abus de notation, les éléments de F_0 sont parfois appelés ensembles négligeables d'observateurs même quand on ne définit pas de mesure sur $(0, F)$. Notons le résultat suivant qui éclaire le sens à donner à F_0 :

I.2.7.- PROPOSITION.

Supposons que pour tout observateur $\xi \in 0$ la partie $\{\xi\} \in F$; alors la classe F_0 est égale à $\{\emptyset\}$ si et seulement si tous les observateurs $\xi \in 0$ sont non-négligeables (c'est-à-dire, si pour tout $\xi \in 0$, il existe $A \in S$ tel que $J(A, \{\xi\}) > 0$)

Sans hypothèse sur F on a plus généralement que la classe F_0 est égale à $\{\emptyset\}$ si et seulement si tous les observateurs $\xi \in O$ sont non négligeables dans le sens suivant : pour tout $\xi \in O$ et tout $F \in F$ tel que $\xi \in F$, il existe $A \in S$ tel que $J(A, F) > 0$.

Le système d'axiomes ci-dessous, que nous avons proposé avec G. COMYN dans notre note [7] sur les "Applications de l'information généralisée à l'analyse des données statistiques", est équivalent aux définitions I.2.2. et I.2.6. et nous montre que l'application J que nous venons de définir possède les propriétés souhaitées au paragraphe 1.2.1.

I.2.8.- L'information apportée par un groupe d'observateurs $F \in F$ sur un événement $A \in S$ est une application J de $S \times F$ dans $\mathbb{R}^+$ vérifiant les trois axiomes suivants :

AXIOME 1 : Pour tout $(A, F) \in S \times F$: $J(A, F) \geq 0$; de plus pour tout $A \in S$ et $F_0 \in F_0$: $J(A, F_0) = 0$.
 F_0 est la classe d'éléments de F contenant $\emptyset$ et qui est définie en I.2.6.

AXIOME 2 : Pour tout $F \in F - F_0$, l'application de S dans $\mathbb{R}^+$ définie par $J(\cdot, F)$ est une information généralisée au sens du paragraphe I.1.1. sur (Ω, S)

C'est-à-dire que si A_1 et A_2 sont deux événements de S tels que

$$\Omega \supseteq A_2 \supseteq A_1 \supseteq \emptyset$$

il vient pour $F \in F - F_0$:

I.2.9.- $0 = J(\Omega, F) \leq J(A_2, F) \leq J(A_1, F) \leq J(\emptyset, F) = +\infty$

AXIOME 3 : Pour tout $A \in S$ fixé, l'application de F dans $\mathbb{R}^+$ définie par $J(A, \cdot)$ est croissante pour l'inclusion.

Soient F_1 et F_2 deux groupes d'observateurs de F tels que

$$\emptyset \subseteq F_1 \subseteq F_2 \subseteq F$$

il vient :

I.2.10.-
$$0 = J(A, \emptyset) \leq J(A, F_1) \leq J(A, F_2) \leq J(A, F) \leq +\infty$$

Les relations I.2.9 et I.2.10. sont des conséquences de la définition I.2.6. appliquée aux relations des propositions I.2.5. et I.2.4.

Dans la suite, pour simplifier les énoncés, nous utilisons la lettre J et nous parlons simplement de l'information sur $(\Omega \times \emptyset, S \times F)$ pour définir une information vérifiant les définitions I.2.2. et I.2.6 ; nous employons la lettre I et nous précisons information sur (Ω, S) quand nous faisons allusion aux définitions de I.1.1.

Pour simplifier l'exposé et pour employer un formalisme classique, nous avons supposé que les espaces mesurables (Ω, S) et $(\emptyset, F)$ sont formés respectivement par des σ -algèbres de boole S et F de parties de Ω et $\emptyset$. Il est clair que l'on peut énoncer les définitions précédentes sous des hypothèses moins restrictives pour S et F . Il convient alors de prendre la précaution de préciser, dans certains énoncés, que les éléments utilisés appartiennent à S ou F . De même, par la suite, nous supposerons toujours a priori que les ensembles de parties sont des σ -algèbres de boole ; mais nous nous réservons la possibilité d'utiliser le cas échéant des structures moins restrictives en précisant à chaque fois le domaine sur lequel nous travaillons. On trouvera toutes les précisions utiles concernant ce type de généralisation dans le paragraphe II.4.

I.3.- DEPENDANCE ET INDEPENDANCE.

L'indépendance de deux éléments de $S \times F$ au sens où nous l'avons définie dans le paragraphe I.2.3. est l'application immédiate de la définition de l'indépendance entre deux éléments d'un treillis (I.1.2.). Comme dans le cadre classique, nous avons besoin, pour les développements ultérieurs, de généraliser cette notion à l'indépendance de sous-algèbres de S et de famille d'observateurs de F .

Dans la théorie de l'information généralisée au sens de J. KAMPE DE FERIET, (I.1.1. : Axiome 3 et Définition a), il est naturel de considérer que c'est l'indépendance des algèbres :

$$A_1 = \{\emptyset, A_1, A_1^c, \Omega\} \quad \text{et} \quad A_2 = \{\emptyset, A_2, A_2^c, \Omega\}$$

qui définit l'indépendance des événements A_1 et A_2 de S .

Comme, pour tout groupe $F \in \mathcal{F}$ d'observateurs fixé, nous supposons que l'application $J(\cdot, F)$ est une information généralisée sur (Ω, S) , nous exprimerons donc aussi l'indépendance des événements A_1 et A_2 de S pour le groupe $F \in \mathcal{F}$ d'observateurs par l'indépendance des algèbres A_1 et A_2 :

$$A' \in A_1, A'' \in A_2 \implies J((A' \cap A''), F) = J(A', F) + J(A'', F)$$

D'une manière générale, si $\{A_i | i \in K\}$ est une famille de sous σ -algèbres de Boole de S , K étant un ensemble quelconque d'indices, nous poserons la définition suivante :

I.3.1.- DEFINITION.

Soit une information J sur $(\Omega \times \mathcal{O}, S \times F)$ et un groupe $F \in \mathcal{F}$ fixé d'observateurs ; nous dirons que les sous σ -algèbres de boole de la famille $\{A_i | i \in K\}$ sont indépendantes en information (respectivement σ -indépendantes en information) pour le groupe F d'observateurs si pour toute sous-famille finie (resp. dénombrable) d'indices $K' \subset K$, on a :

$$J\left(\bigcap_{i \in K'} A_i, F\right) = \sum_{i \in K'} J(A_i, F)$$

quel que soient les événements $A_i \in \mathcal{A}_i$.

REMARQUE SUR LA NOTION D'INDEPENDANCE.

De manière symétrique, on pourrait penser généraliser la définition de l'indépendance de deux groupes d'observateurs F_1 et F_2 de F pour un événement A fixé, en exigeant l'indépendance des sous-tribus F_1 et F_2 :

$$F_i = \{\emptyset, F_i, F_i^c, \Omega\} \quad i = 1, 2$$

Nous ne le faisons pas car cette définition aboutit à des résultats contradictoires avec le sens intuitif que nous attachons à $J(A, F)$. Du fait que F_1 et F_2 sont indépendants pour l'observation de A il n'en résulte nullement que F_1^c et F_2^c sont indépendants.

Nous venons de définir l'indépendance et la σ -indépendance d'une famille de sous-tribus de S pour le même groupe d'observateurs. Nous allons maintenant définir l'indépendance et la σ -indépendance d'une famille de groupes d'observateurs pour l'observation d'un même élément $A \in S$ ou d'une sous-tribu A de S . Soit $\{F_i | i \in K, F_i \in F\}$ une famille de groupes d'observateurs F -mesurable, K étant un ensemble quelconque d'indices.

I.3.2.- DEFINITION.

Soit une information J sur $(\Omega \times \Omega, S \times F)$ et un élément $A \in S$ fixé (respectivement une sous-tribu A de S fixée) ; nous dirons que les groupes d'observateurs F -mesurables de la famille $\{F_i | i \in K\}$ sont indépendants pour l'observation de l'événement A (resp. des événements de la sous-tribu A) si pour toute sous-famille finie d'indices $K' \subset K$ on a pour A (resp. $\forall A \in A$) :

$$J(A, \bigcup_{i \in K'} F_i) = \sum_{i \in K'} J(A, F_i)$$

Les groupes d'observateurs seront dits σ -indépendants si la relation est vraie pour toute sous-famille finie ou dénombrable d'indices $K' \subset K$.

Dans la suite, sous certaines hypothèses, nous serons amenés à lier certains groupes d'observateurs $F_i \in F$ à l'observation de sous-tribus particulières $A_i \in S$. Soit $\{(A_i, F_i) | i \in K\}$ une famille de couples formée par une sous-tribu A_i de S et un groupe d'observateurs $F_i \in F$, K étant un ensemble quelconque d'indices.

I.3.3.- DEFINITION.

Soit une information J sur $(\Omega \times 0, S \times F)$; nous dirons que les groupes d'observateurs $\{F_i | i \in K\}$ sont indépendants pour l'observation des tribus $\{A_i | i \in K\}$ si, pour toute sous-famille finie d'indices $K' \subset K$ on a :

$$J\left(\left(\bigcap_{i \in K'} A_i\right), \left(\bigcup_{i \in K'} F_i\right)\right) = \sum_{i \in K'} J(A_i, F_i)$$

quels que soient les événements $A_i \in A_i$.

Nous parlerons de σ -indépendance si la relation est vraie pour toute sous-famille finie ou dénombrable d'indices $K' \subset K$.

Si $F_i = F$ pour tout $i \in K$ on retrouve la définition I.3.1.; et si $A_i = A$ pour tout $i \in K$, on retrouve la définition I.3.2.

La définition de l'information conditionnelle est une simple application de la définition générale, sur le treillis engendré sur $S \times F$ par la relation d'ordre .

I.3.4.- DEFINITION.

L'information conditionnelle fournie par l'élément $(A_1, F_1) \in S \times F$ connaissant l'information apportée par $(A_2, F_2) \in S \times F$ est définie par :

$$J((A_1, F_1) / (A_2, F_2)) = J((A_1, F_1) \wedge (A_2, F_2)) - J(A_2, F_2)$$

$$\text{si } J(A_2, F_2) < +\infty$$

c'est-à-dire :

$$J((A_1, F_1) / (A_2, F_2)) = J((A_1 \cap A_2), (F_1 \cup F_2)) - J(A_2, F_2)$$

$$\text{si } J(A_2, F_2) < +\infty$$

Si $J(A_2, F_2) = +\infty$ cela implique $J((A_1 \cap A_2), (F_1 \cup F_2)) = +\infty$ et le second membre est indéterminé. On peut, le cas échéant, lever l'indétermination en posant :

$$J((A_1, F_1) / (A_2, F_2)) = J(A_1, F_1) \text{ si } J(A_2, F_2) = +\infty$$

Ce choix est le seul qui rend équivalentes les deux propositions :

" (A_1, F_1) est indépendant de (A_2, F_2) " et

" $J((A_1, F_1) / (A_2, F_2)) = J(A_1, F_1)$ " .

Pour $(A_2, F_2) \in S \times F$ fixé, l'application $J(. / (A_2, F_2))$ définie sur $S \times F$, à valeurs dans $\mathbb{R}^+$, vérifie les hypothèses de la définition I.2.2. ; $J(. / (A_2, F_2))$ définit donc en ce sens une information sur $(\Omega \times 0, S \times F)$ ce qui justifie la terminologie employée.

Nous avons évidemment les relations suivantes quand les expressions sont définies :

$$\begin{aligned} J((A_1, F_1) / (A_2, F_2)) &= J((A_1 \cap A_2), (F_1 \cup F_2)) / (A_2, F_2) \\ &= J((A_1 \cap A_2), F_1) / (A_2, F_2) \\ &= J((A_1, (F_1 \cup F_2)) / (A_2, F_2)) \\ &= J((A_1, (F_1 \cup F_2)) / (A_2, (F_1 \cup F_2))) \\ &\quad + J((A_2, F_1) / (A_2, F_2)) \\ &= J(((A_1 \cap A_2), (F_1 \cup F_2)) / (A_2, (F_1 \cup F_2))) \\ &\quad + J((A_2, (F_1 \cup F_2)) / (A_2, F_2)) \\ &= J(((A_1 \cap A_2), (F_1 \cup F_2)) / ((A_1 \cap A_2), F_2)) \\ &\quad + J(((A_1 \cap A_2), F_2) / (A_2, F_2)) \end{aligned}$$

On peut donner de $J((A_1, F_1) / (A_2, F_2))$ l'interprétation suivante : Information apportée par les observateurs de $F_1 \cup F_2$ sur la réalisation de l'événement A_1 (ou $A_1 \cap A_2$) connaissant l'information apportée par les observateurs de F_2 sur la réalisation de l'événement A_2 .

On a nécessairement :

$$J((\Omega, \emptyset) / (A_2, F_2)) = 0$$

$$J((\emptyset, \emptyset) / (A_2, F_2)) = +\infty$$

$$J((\emptyset, F_1) / (A_2, F_2)) = +\infty \text{ si } F_1 \cup F_2 \notin \mathcal{F}_0$$

mais on peut obtenir aussi :

$$J((A_1, \emptyset) / (A_2, F_2)) = J((A_1, F_2) / (A_2, F_2)) > 0$$

$$J((\Omega, F_1) / (A_2, F_2)) = J((A_2, F_1) / (A_2, F_2)) > 0$$

ce qui implique, en particulier, que $J(. / (A_2, F_2))$ ne vérifie pas toutes les hypothèses de la définition I.2.6.

La définition suivante généralise la notion d'information mutuelle apportée par deux événements.

I.3.5.- DEFINITION.

On appelle information mutuelle entre les éléments

$(A_i, F_i) \in \mathcal{S} \times \mathcal{F}$, $i = 1, 2$ *la quantité*

$$J((A_1, F_1) ; (A_2, F_2)) = J(A_1, F_1) + J(A_2, F_2) - J((A_1, F_1) \wedge (A_2, F_2))$$

$$\text{si} \quad J(A_i, F_i) < +\infty, \quad i = 1, 2$$

c'est-à-dire :

$$J((A_1, F_1) ; (A_2, F_2)) = J(A_1, F_1) + J(A_2, F_2) - J((A_1 \cap A_2), (F_1 \cup F_2))$$

$$\text{si} \quad J(A_i, F_i) < +\infty, \quad i = 1, 2$$

Si l'une des informations $J(A_i, F_i)$, $i = 1, 2$ est égale à $+\infty$ cela implique $J((A_1, F_1) \wedge (A_2, F_2)) = +\infty$ et le second membre est indéterminé ; on peut, le cas échéant, lever l'indétermination en posant :

$$J((A_1, F_1) ; (A_2, F_2)) = 0$$

Ce choix est le seul qui rend équivalentes les deux propositions :

" (A_1, F_1) est indépendant de (A_2, F_2) " et

" $J((A_1, F_1) ; (A_2, F_2)) = 0$ "

Quand les expressions sont définies, nous avons les relations :

$$J((A_1, F_1) ; (A_2, F_2)) = J(A_1, F_1) - J((A_1, F_1) / (A_2, F_2))$$

$$J((A_1, F_1) ; (A_2, F_2)) = J(A_2, F_2) - J((A_2, F_2) / (A_1, F_1))$$

Notons enfin que l'information mutuelle n'est pas une information ; la quantité $J(.,.)$ n'est pas obligatoirement positive : $J(.,.) \in [-\infty, +\infty[$, et n'a pas de propriété de monotonie.

CHAPITRE II

INFORMATION SUR UN

ESPACE PRODUIT

La formalisation des exemples présentés dans l'introduction de ce mémoire nous amène naturellement à orienter notre étude vers les espaces produits. Notons Ω un groupe de personnes soumises à une enquête et O l'ensemble des questions de cette enquête. Définissons, pour chaque question $\xi \in O$, l'ensemble Ω_ξ des modalités du caractère étudié et S_ξ une tribu de parties de Ω_ξ . L'espace des observations $(\hat{\Omega}, \hat{S}) = (\prod_{\xi \in O} \Omega_\xi, \otimes_{\xi \in O} S_\xi)$, produit des espaces mesurables $\{(\Omega_\xi, S_\xi) \mid \xi \in O\}$, est défini dès la conception du questionnaire de l'enquête. C'est l'application T associant à chaque individu $\omega \in \Omega$ ses réponses $\{T_\xi(\omega) \in \Omega_\xi \mid \xi \in O\}$ au questionnaire, qui formalise la réalisation concrète de l'enquête ; elle nous permet de passer des informations sur l'espace des réponses possibles $(\hat{\Omega}, \hat{S})$ aux informations effectivement obtenues sur l'espace (Ω, S) des personnes interrogées.

L'objet de ce chapitre est d'étudier les liens existant entre une information J apportée par un groupe d'observateurs de (O, F) sur un événement de $(\hat{\Omega}, \hat{S})$ et une information généralisée I sur $(\hat{\Omega}, \hat{S})$ au sens de J. KAMPE DE FERIET et B. FORTE, quand $(\hat{\Omega}, \hat{S})$ est un espace produit indicé par l'ensemble des observateurs O , espace que nous appelons espace des observations.

Ce travail nous procure des techniques de construction des informations J à partir des informations I au sens de J. KAMPE DE FERIET, pour les espaces produits de ce type et inversement.

Nous nous sommes placés dans un cadre très général : l'ensemble O et les espaces mesurables $\{(\Omega_\xi, S_\xi) | \xi \in O\}$ sont quelconques, finis ou non ; les notations et les démonstrations sont donc adaptées aux produits infinis d'espaces mesurables. Dans les applications du type de celles proposées dans nos exemples, ce sont essentiellement des produits finis d'espaces mesurables qui interviennent. On trouvera dans les chapitres suivants diverses applications des résultats de cette partie quand les ensembles $\{\Omega_\xi | \xi \in O\}$ sont finis et quand l'ensemble O est lui aussi formé d'un nombre fini d'observateurs.

II.1. - INFORMATIONS SUR $(\hat{\Omega} \times O, \hat{S} \times F)$.

Soit une famille quelconque d'espaces mesurables $\{(\Omega_\xi, S_\xi) | \xi \in O\}$ indicés par O . Conformément aux notations et conventions adaptées précédemment (I.1.3.d) nous notons $(\hat{\Omega}, \hat{S})$ l'espace mesurable produit des espaces $\{(\Omega_\xi, S_\xi) | \xi \in O\}$.

$$\hat{\Omega} = \prod_{O} \Omega_\xi$$

$$\hat{S} = \otimes_{O} S_\xi$$

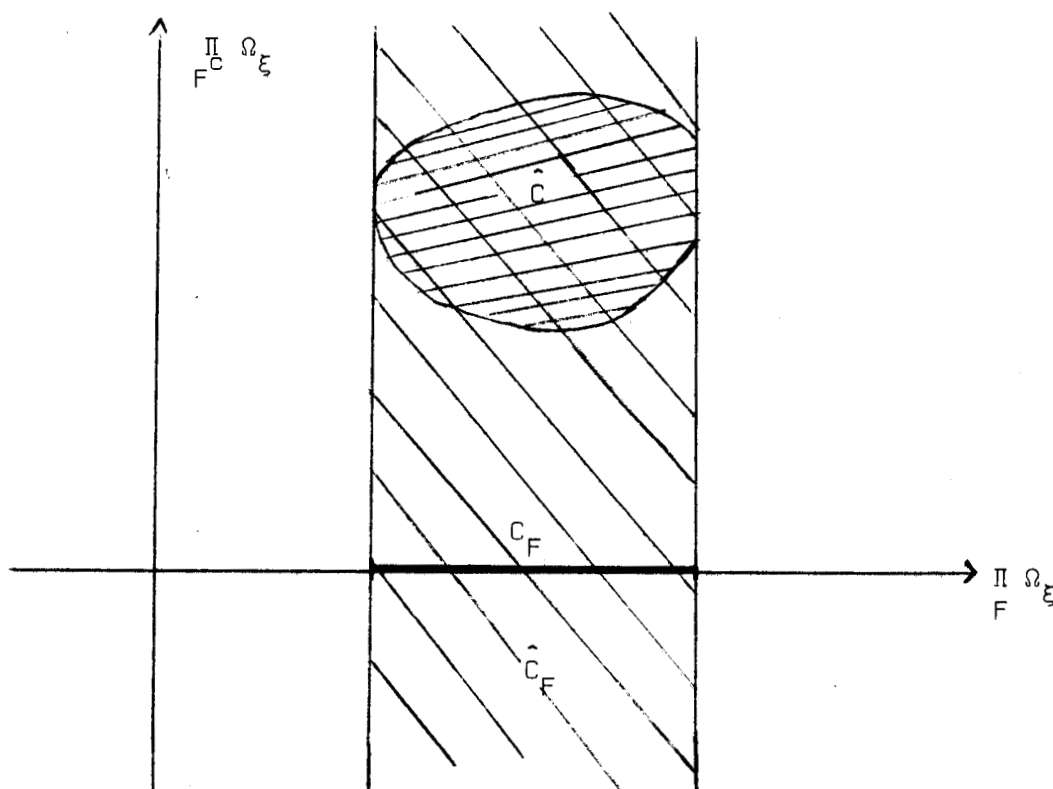
Soit I une information sur $(\hat{\Omega}, \hat{S})$ qui vérifie les hypothèses du paragraphe I.1.1.

Essayons d'exprimer l'idée des propositions suivantes par un graphique.

Soit $F \in \mathcal{F}$ un groupe d'observateurs de $\mathcal{O}$; représentons l'espace $(\hat{\Omega}, \hat{S})$ comme le produit des espaces $(\Pi_F \Omega_\xi, \otimes_F S_\xi)$ et $(\Pi_{F^c} \Omega_\xi, \otimes_{F^c} S_\xi)$. Notons C_F la projection sur $\Pi_F \Omega_\xi$ d'un événement quelconque $\hat{C} \in \hat{S}$; par convention, nous avons convenu de noter $\hat{C}_F$ le cylindre de $\hat{S}$ de base C_F .

Pour définir l'information $J(\hat{C}, F)$ apportée par l'observation de la réalisation de l'événement $\hat{C}$ par les observateurs du groupe F nous posons :

$$J(\hat{C}, F) = I(\hat{C}_F).$$



C'est-à-dire que $J(\hat{C}, F)$ est égale à l'information généralisée apportée par le plus petit événement $\hat{\otimes}_F S_\xi$ -mesurable qui contient $\hat{C}$. Intuitivement

les observateurs du groupe F ne perçoivent pas l'événement $\hat{C}$ de l'espace total $(\hat{\Omega}, \hat{S})$ mais seulement la trace C_F de cet événement dans leur espace propre $(\prod_F \Omega_\xi, \otimes_F S_\xi)$.

II.1.1. - HYPOTHESE (I.1.3.e).

Dans la suite nous supposons que pour tout $(\hat{C}, F) \in \hat{S} \times F$, $F \neq \emptyset$, la projection C_F de $\hat{C}$ sur $\prod_F \Omega_\xi$ est $\otimes_F S_\xi$ -mesurable, ou, ce qui est équivalent, que le cylindre associé $\hat{C}_F = C_F \times \prod_{F^c} \Omega_\xi$, est $\hat{S}$ -mesurable (c'est-à-dire aussi $\otimes_F S_\xi$ -mesurable).

II.1.2.- PROPOSITION.

Soit une information I , au sens de I.1.1., définie sur $(\hat{\Omega}, \hat{S})$
l'application J définie pour tout $(\hat{C}, F) \in \hat{S} \times F$ par

$$J(\hat{C}, F) = I(\hat{C}_F) \quad \text{si } F \neq \emptyset$$

$$J(\hat{C}, F) = 0 \quad \text{si } F = \emptyset$$

est une information J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ au sens de I.2.2. et I.2.6.
avec $F_0 = \{\emptyset\}$.

Montrons que toute information I permet de définir une information J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ de la manière indiquée. Pour tout couple $(\hat{C}_i \times F_i) \in \hat{S} \times F$, $F_i \neq \emptyset$, $i = 1, 2$, on a (I.1.3.d) :

$$\left[\begin{array}{l} \hat{C}_1 \subseteq \hat{C}_2 \\ F_1 \supseteq F_2 \end{array} \right] \Rightarrow \widehat{[C_1]}_{F_1} \subseteq \widehat{[C_2]}_{F_2}$$

en notant $\widehat{C_i}_{F_i}$, ($i = 1, 2$), le cylindre, supposé $\hat{S}$ -mesurable dont la base est la projection canonique de $\hat{C}_i$ sur $\Pi_{F_i} \Omega_{\xi}$. La démonstration est immédiate :

$$\widehat{C_1}_{F_1} \subseteq \widehat{C_2}_{F_2} \Rightarrow I(\widehat{C_1}_{F_1}) \geq I(\widehat{C_2}_{F_2}) \geq 0$$

et

$$J(\hat{C}_1, F_1) \geq J(\hat{C}_2, F_2) \geq 0.$$

Les hypothèses de I.2.2. sont alors vérifiées. De plus, il est clair que, pour tout $F \in \mathcal{F}$ et $F \neq \emptyset$:

$$J(\emptyset, F) = I(\emptyset) = +\infty$$

c'est-à-dire que la classe F_0 des groupes d'observateurs qui n'apportent pas d'information est réduite à l'ensemble vide : $F_0 = \{\emptyset\}$. On obtient effectivement les valeurs universelles de I.2.6. ; pour tout $F \in \mathcal{F} - F_0$

- . $J(\hat{\Omega}, F) = I(\hat{\Omega}) = 0$ et $J(\emptyset, F) = I(\emptyset) = +\infty$
- . $J(\hat{\Omega}, \emptyset) = J(\emptyset, \emptyset) = 0$

La proposition suivante est une conséquence immédiate de la proposition II.1.2.

II.1.3.- PROPOSITION.

Les informations I et J de la proposition II.1.2. vérifient les relations suivantes ; pour tout $(\hat{C}, F) \in \hat{S} \times \mathcal{F}$, $F \neq \emptyset$:

- . $I(\hat{C}_F) = J(\hat{C}, F) = J(\hat{C}_F, \emptyset)$
- . $I(\hat{C}) = J(\hat{C}, \emptyset)$.

Inversement, étant donné une information J' au sens I.2.2.-I.2.6. sur $(\hat{\Omega} \times 0, \hat{S} \times \mathcal{F})$ avec $F_0 = \{\emptyset\}$, nous savons par définition de J' que pour tout $F \in \mathcal{F} - F_0$ fixé, les applications I_F définies sur $\hat{S}$ par

$I_F(.) = J'(. , F)$ sont des informations sur $(\hat{\Omega}, \hat{S})$ au sens de J. KAMPE DE FERIET
 En particulier, nous pouvons donc toujours définir sur $(\hat{\Omega}, \hat{S})$ une information I
 en posant $I(.) = J'(. , 0)$. Appliquons la proposition II.1.2. à cette informa-
 tion, pour tout $(\hat{C} , F) \in \hat{S} \times F$ et $F \neq \emptyset$:

$$J(\hat{C} , F) = I(\hat{C}_F) = J'(\hat{C}_F , 0)$$

de plus, il est clair que si : $J(\hat{C} , F) = J(\hat{C}_F , 0) = J'(\hat{C}_F , 0)$, en général,
 nous avons : $J'(\hat{C} , F) \neq J'(\hat{C}_F , 0)$, ce qui entraîne $J'(\hat{C} , F) \neq J(\hat{C} , F)$.
 Etant donné une information J' sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ cette information n'est
 pas nécessairement liée à l'information $I(.) = J'(. , 0)$ sur $(\hat{\Omega}, \hat{S})$ par les
 relations de la proposition II.1.2. Nous avons le résultat suivant :

II.1.4.- THEOREME.

Etant donné une information J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ au sens I.2.2.-I.2.6.
 avec $F_0 = \{\emptyset\}$, définissons une information I sur $(\hat{\Omega}, \hat{S})$ au sens de
 I.1.1. par $I(.) = J(. , 0)$.

Une condition nécessaire suffisante pour que les informations I et J
 vérifient les relations de la proposition II.1.2. est que pour tout $(\hat{C} , F) \in \hat{S} \times F$,
 $F \neq \emptyset$:

$$J(\hat{C} , F) = J(\hat{C}_F , 0).$$

La condition nécessaire résulte de la proposition II.1.3. et la
 condition suffisante du raisonnement précédent.

Notons que la condition $J(\hat{C} , F) = J(\hat{C}_F , 0)$ implique en particulier
 pour tout $F \neq \emptyset$:

$$J(\emptyset , F) = J(\emptyset , 0) = + \infty$$

c'est-à-dire $F_0 = \{\emptyset\}$. On pourrait cependant généraliser les propositions

II.1.2., II.1.3. et II.1.4. en supposant $F_0 \neq \{\emptyset\}$. Cela reviendrait pratiquement à remplacer, dans les énoncés précédents, les relations $F \neq \emptyset$ par $F \in F_0$ et $F = \emptyset$ par $F \notin F_0$.

REMARQUE SUR LES NOTIONS DE DEPENDANCE ET D'INDEPENDANCE.

Nous supposons que la condition du théorème II.1.4. est vérifiée, c'est-à-dire que nous nous donnons une information I sur $(\hat{\Omega}, \hat{S})$ et nous construisons l'information J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ à partir de I suivant la technique indiquée dans la proposition II.1.2.

Nous allons étudier rapidement les relations entre les notions d'indépendance et de dépendance pour les informations I et J . Pour éviter des répétitions inutiles nous supposons toujours implicitement que les groupes d'observateurs de F utilisés : $F, F_1, F_2, \dots$ sont différents de l'ensemble vide.

On sait (I.1.1.) que l'information fournie par la réalisation simultanée de deux événements $\hat{C}_1$ et $\hat{C}_2$ de $\hat{S}$ indépendants pour l'information I est la somme des informations fournies par chacun d'eux :

$$I(\hat{C}_1 \cap \hat{C}_2) = I(\hat{C}_1) + I(\hat{C}_2).$$

Suivant notre définition I.2.3., deux éléments $(\hat{C}_1, F_1) \in \hat{S} \times F$, $1 = 1, 2$ sont J -indépendants en information si :

$$J((\hat{C}_1 \cap \hat{C}_2), (F_1 \cup F_2)) = J(\hat{C}_1, F_1) + J(\hat{C}_2, F_2).$$

Le cylindre $\overbrace{(C_1 \cap C_2)}^{F_1 \cup F_2}$ dont la base est la projection de

l'ensemble $\hat{C}_1 \cap \hat{C}_2$ sur l'espace $\prod_{F_1 \cup F_2} \Omega_{\xi}$ est inclus dans l'intersection des cylindres $\overbrace{(C_1)}^{F_1}$ et $\overbrace{(C_2)}^{F_2}$ et en règle générale

$\overbrace{(C_1 \cap C_2)}^{F_1 \cup F_2} \neq \overbrace{(C_1)}^{F_1} \cap \overbrace{(C_2)}^{F_2}$. La J -indépendance en information

des éléments $(\hat{C}_i, F_i)$, $i = 1, 2$ qui s'écrit aussi :

$$I(\widehat{(C_1 \cap C_2)}_{F_1 \cup F_2}) = I(\widehat{(C_1)}_{F_1}) + I(\widehat{(C_2)}_{F_2})$$

n'implique donc pas la I-indépendance des cylindres $\widehat{(C_i)}_{F_i}$, $i = 1, 2$ qui s'écrit :

$$I(\widehat{(C_1)}_{F_1} \cap \widehat{(C_2)}_{F_2}) = I(\widehat{(C_1)}_{F_1}) + I(\widehat{(C_2)}_{F_2}).$$

De même la I-indépendance des cylindres $\widehat{(C_i)}_{F_i}$, $i = 1, 2$ n'implique par en général la J-indépendance des $(\hat{C}_i, F_i)$, $i = 1, 2$.

Notons cependant le résultat suivant : soient $F_i \in \mathcal{F}$, $i = 1, 2$ deux groupes d'observateurs donnés; pour tout $\hat{C}_i \in \hat{\mathcal{C}}_{S_{F_i}}$, $i = 1, 2$ c'est-à-dire pour tout couple de cylindres de $\hat{S}$ à base respectivement dans $\Pi_{F_1} \Omega_{F_1}$ et $\Pi_{F_2} \Omega_{F_2}$ la I-indépendance en information des événements $\hat{C}_1, \hat{C}_2$ implique la J-indépendance des éléments $(\hat{C}_1, F_1), (\hat{C}_2, F_2)$ et inversement.

En effet, dans ce cas : $\widehat{(C_i)}_{F_i} = \hat{C}_i$, $i = 1, 2$ et

$$\hat{C}_1 \cap \hat{C}_2 = \widehat{(C_1)}_{F_1} \cap \widehat{(C_2)}_{F_2} = \widehat{(C_1 \cap C_2)}_{F_1 \cup F_2}.$$

Ce qui nous permet d'écrire le résultat demandé :

$$\begin{aligned} J(\hat{C}_1 \cap \hat{C}_2, (F_1 \cup F_2)) &= J(\hat{C}_1, F_1) + J(\hat{C}_2, F_2) \\ &\iff I(\widehat{(C_1 \cap C_2)}_{F_1 \cup F_2}) = I(\widehat{(C_1)}_{F_1}) + I(\widehat{(C_2)}_{F_2}) \\ &\iff I(\widehat{(C_1)}_{F_1} \cap \widehat{(C_2)}_{F_2}) = I(\widehat{(C_1)}_{F_1}) + I(\widehat{(C_2)}_{F_2}) \end{aligned}$$

notons que dans ce cas : $J(\hat{C}_i, F_i) = I(\widehat{(C_i)}_{F_i}) = J(\hat{C}_i, \emptyset)$ $i = 1, 2$.

Notons aussi qu'étant donné une famille quelconque de groupes d'observateurs de $F : \{F_i | i \in K\}$, il est évident que la J -indépendance (respectivement la J - σ -indépendance) des groupes d'observateurs $\{F_i | i \in K\}$ pour l'observation des sous- σ -algèbres de Boole $\{\hat{\otimes}_{F_i} S_\xi | i \in K\}$ telle qu'elle est définie en I.3.3., implique la I -indépendance (resp. la I - σ -indépendance) des sous- σ -algèbres de Boole $\{\hat{\otimes}_{F_i} S_\xi | i \in K\}$ et inversement.

Enfin soit $(\hat{C}_i, F_i) \in \hat{S} \times F$, $i = 1, 2$

on aura, en règle générale :

$$J((\hat{C}_1, F_1)/(\hat{C}_2, F_2)) \neq I((\hat{C}_1)_{F_1}/(\hat{C}_2)_{F_2}) .$$

Cependant, comme précédemment, nous avons le résultat suivant :

soient $F_i \in F$, $i = 1, 2$ deux groupes d'observateurs donnés ; pour tout $\hat{C}_i \in \hat{\otimes}_{F_i} S_\xi$, $i = 1, 2$, nous avons :

$$J((\hat{C}_1, F_1)/(\hat{C}_2, F_2)) = I((\hat{C}_1)_{F_1}/(\hat{C}_2)_{F_2}) \text{ si } J(\hat{C}_2, F_2) = I((\hat{C}_2)_{F_2}) < +\infty$$

ce qui s'écrit aussi :

$$J((\hat{C}_1, F_1)/(\hat{C}_2, F_2)) = J((\hat{C}_1, 0)/(\hat{C}_2, 0)) = I(\hat{C}_1/\hat{C}_2) \text{ si } I(\hat{C}_2) < +\infty$$

II.2.- APPLICATIONS.

Soit T une application mesurable de (Ω, S) dans $(\hat{\Omega}, \hat{S})$ telle que $T(\Omega) \in \hat{S}$. Cette variable aléatoire permet, par exemple, d'associer à chaque individu ω d'une population statistique Ω le point $T(\omega) \in \hat{\Omega}$ représentatif des caractéristiques de cet individu. Notons A_T la sous-tribu de S image

inverse par T de la σ -algèbre de Boole $\hat{S}$ de parties de $\hat{\Omega}$: $A_T = T^{-1}(\hat{S})$ et notons respectivement $\{A_\xi | \xi \in O\}$ les images inverses des sous- σ -algèbres de Boole de parties $\{\hat{S}_\xi | \xi \in O\}$. Avec les conventions précédemment adoptées, on peut considérer $A_\xi = T^{-1}(\hat{S}_\xi)$ comme la sous-tribu de A_T engendrée par l'image inverse par T de la sous-tribu des cylindres de $\hat{S}$ à base dans Ω_ξ ; et aussi $A_\xi = T_\xi^{-1}(S_\xi)$ comme la sous-tribu de A_T engendrée par l'image inverse, par l'application coordonnée T_ξ de T , de la tribu S_ξ de Ω_ξ . Il est clair que la tribu A_T est engendrée par la famille $\{A_\xi | \xi \in O\}$. Notons que si les sous-tribus de $\hat{S}$ de la famille $\{\hat{S}_\xi | \xi \in O\}$ sont σ -M-indépendantes, les sous-tribus de S de la famille $\{A_\xi | \xi \in O\}$ ne sont pas nécessairement M-indépendantes (sauf si $T(\Omega) = \hat{\Omega}$).

L'application T permet de définir sur $(\Omega \times O, A_T \times F)$ une information J_1 qui vérifie les définitions I.2.2. et I.2.6. avec $F_0 = \{\emptyset\}$ en posant, pour tout $(A, F) \in A_T \times F$

$$J_1(A, F) = J(T(A), F).$$

J étant une information quelconque sur $(\hat{\Omega} \times O, \hat{S} \times F)$.

Notons que, pour tout $A \in A_T$: $T(A) \in \hat{S}$; en effet, $A \in A_T$ implique qu'il existe $\hat{C} \in \hat{S}$ tel que $A = T^{-1}(\hat{C})$ et comme $T(\Omega) \in \hat{S}$, nous avons $T(A) = T(\Omega) \cap \hat{C} \in \hat{S}$.

Nous pouvons aussi définir une information I_1 sur (Ω, A_T) en posant, pour tout $A \in A_T$:

$$I_1(A) = I(T(A)).$$

I étant une information quelconque sur $(\hat{\Omega}, \hat{S})$. Si de plus I et J vérifient les relations de la proposition II.1.2. ,

il est clair que, pour tout $(A, F) \in A_T \times F$ on a :

$$J_1(A, F) = I_1(A_F) \quad \text{si } F \neq \emptyset$$

$$J_1(A, F) = 0 \quad \text{si } F = \emptyset$$

avec $A_F = T^{-1}(T(A)_F)$, en notant $T(A)_F$ le cylindre dont la base est la projection de $T(A)$ sur l'espace $\prod_F \Omega_\xi$. Inversement, pour tout $A \in A_T$

$$I_1(A) = J_1(A, \emptyset).$$

II.3. - INFORMATIONS SUR $(\prod_{O_1} \Omega_\xi \times O_2, \otimes_{O_1} S_\xi \times F_2)$.

Les résultats des paragraphes précédents sont liés essentiellement au fait que l'espace des événements est un espace produit $(\hat{\Omega}, \hat{S}) = (\prod_{O_1} \Omega_\xi, \otimes_{O_1} S_\xi)$ et que l'on utilise l'ensemble O des observateurs comme ensemble des indices. On peut imaginer que l'espace des événements est un espace produit indicé par un ensemble $O_1 : (\prod_{O_1} \Omega_\xi, \otimes_{O_1} S_\xi)$ et que l'espace des observateurs (O_2, F_2) est formé par un ensemble O_2 a priori distinct de O_1 . On peut aussi supposer que $O_1 = O_2$ mais que l'information J_2 que l'on se donne a priori sur $(\prod_{O_1} \Omega_\xi \times O_2, \otimes_{O_1} S_\xi \times F_2)$ ne vérifie pas les hypothèses du théorème II.1.4. Ces deux remarques nous conduisent à proposer une généralisation des résultats du paragraphe II.1. Notons cependant que nous n'utiliserons pas directement ces nouveaux résultats dans les chapitres ultérieurs.

Etant données J_2 une information sur $(\prod_{O_1} \Omega_\xi \times O_2, \otimes_{O_1} S_\xi \times F_2)$ au sens I.2.2. - I.2.6. et F_2^0 la classe des groupes d'observateurs de F_2 qui n'apportent pas d'information pour J_2 , on définit sur $(\prod_{O_1} \Omega_\xi, \otimes_{O_1} S_\xi)$ une information I_{O_2} en posant pour tout $C \in \prod_{O_1} S_\xi$

$$I_{0_2}(C) = J_2(C, 0_2).$$

Définissons maintenant une information J_{0_2} au sens I.2.2. - I.2.6. suivant la technique de la proposition II.1.2. ; pour tout $(C, F_1) \in \bigotimes_{0_1} S_\xi \times F_1$ nous avons

$$J_{0_2}(C, F_1) = I_{0_2}(C_{F_1}) = J_2(C_{F_1}, 0_2) \quad \text{si } F_1 \neq \emptyset$$

$$J_{0_2}(C, F_1) = 0 \quad \text{si } F_1 = \emptyset$$

Dans ce paragraphe par un abus de notation qui identifie les cylindres de base dans $\Pi_{F_1} \Omega_\xi$ à leur base, nous noterons aussi $\bigotimes_{F_1} S_\xi$ la sous σ -algèbre de $\bigotimes_{0_1} S_\xi$ des cylindres mesurables à base dans $\Pi_{F_1} \Omega_\xi$ et C_{F_1} ces cylindres.

On pourrait de même, pour tout groupe d'observateurs $F_2 \in F_2 - F_2^0$, définir une information I_{F_2} sur $(\Pi_{0_1} \Omega_\xi, \bigotimes_{0_1} S_\xi)$ par :

$$\forall C \in \bigotimes_{0_1} S_\xi \quad I_{F_2}(C) = J_2(C, F_2)$$

et une information J_{F_2} au sens I.2.2. et I.2.6. sur $(\Pi_{0_1} \Omega_\xi \times 0_1, \bigotimes_{0_1} S_\xi \times F_1)$ par :

$$\forall (C, F_1) \in \bigotimes_{0_1} S_\xi \times F_1 \quad J_{F_2}(C, F_1) = J_2(C_{F_1}, F_2) \quad \text{si } F_1 \neq \emptyset.$$

Associons à chaque couple $(F_1, F_2) \in F_1 \times F_2$ le pavé mesurable $F_1 \times F_2$.

En posant $J(C, (F_1 \times F_2)) = J_2(C_{F_1}, F_2)$ nous définissons une information sur

$(\prod_{0_1} \Omega_{\xi} \times (0_1 \times 0_2), \otimes_{0_1} S_{\xi} \times (F_1 \times F_2))$ par une technique qui généralise celle de la proposition II.1.2.

II.3.1.- PROPOSITION.

Etant données une information J_2 sur $(\prod_{0_1} \Omega_{\xi} \times 0_2, \otimes_{0_1} S_{\xi} \times F_2)$ au sens I.2.2. - I.2.6. et F_2^0 la classe des groupes d'observateurs de F_2 qui n'apportent pas d'information pour J_2 , l'application J définie sur

$(\prod_{0_1} \Omega_{\xi} \times (0_1 \times 0_2), \otimes_{0_1} S_{\xi} \times (F_1 \times F_2))$, pour tout $(C, F_1, F_2) \in \otimes_{0_1} S_{\xi} \times F_1 \times F_2$

par :

$$J(C, (F_1 \times F_2)) = J_2(C_{F_1}, F_2) \text{ si } (F_1, F_2) \in (F_1 \times F_2) - F_{12}^0$$

$$J(C, (F_1 \times F_2)) = 0 \quad (F_1, F_2) \in F_{12}^0$$

$$\text{avec } F_{12}^0 = \{(F_1, F_2) \in F_1 \times F_2 \mid F_1 = \emptyset \text{ ou } F_2 \in F_2^0\}$$

est une information au sens I.2.2. - I.2.6. et F_{12}^0 est la classe des groupes d'observateurs de $F_1 \times F_2$ qui n'apportent pas d'information pour J .

En effet soient deux éléments (C, F_1, F_2) et (C', F'_1, F'_2) de $\otimes_{0_1} S_{\xi} \times F_1 \times F_2$ tels que $C \subset C'$ et $F_1 \times F_2 \supset F'_1 \times F'_2$, ce qui implique $F_1 \supset F'_1, F_2 \supset F'_2$ et $C_{F_1} \subset C'_{F'_1}$. Par définition de J_2 :

$$J(C, (F_1 \times F_2)) = J_2(C_{F_1}, F_2) \geq J_2(C'_{F'_1}, F'_2) = J(C', (F'_1 \times F'_2))$$

quand (F_1, F_2) et (F'_1, F'_2) n'appartiennent pas à F_{12}^0 .

Vérifions que F_{12}^0 possède les propriétés définies en I.2.6.

$$- \emptyset \in F_{12}^0 ; \text{ en effet } \begin{cases} (\emptyset, F_2) \in F_{12}^0 & \text{pour tout } F_2 \in F_2 \\ (F_1, \emptyset) \in F_{12}^0 & \text{pour tout } F_1 \in F_1 \text{ car } \emptyset \in F_2^0 \end{cases}$$

- F_{12}^0 est héréditaire ; si

$$(F_1, F_2) \in F_{12}^0 \text{ et } (F'_1, F'_2) \in F_1 \times F_2 \text{ avec } F_1 \times F_2 \supset F'_1 \times F'_2$$

cela implique :

$$\left[\begin{array}{l} F_1 = \emptyset \text{ et } F_1 \supset F'_1 \Rightarrow F'_1 = \emptyset \\ \text{et/ou } F_2 \in F_2^0 \text{ et } F_2 \supset F'_2 \Rightarrow F'_2 \in F_2^0 \text{ car } F_2^0 \text{ est héréditaire.} \\ \text{c'est-à-dire } (F'_1, F'_2) \in F_{12}^0. \end{array} \right.$$

- L'ensemble $(F_1 \times F_2) - F_{12}^0$ est non vide car il contient au moins l'élément $(0_1, 0_2)$

Il est clair, étant donné les propriétés de F_{12}^0 , que J vérifie les hypothèses des définitions I.2.2. et I.2.6. et en particulier

$$\begin{aligned} (F_1, F_2) \in (F_1 \times F_2) - F_{12}^0 ; \quad J(\pi_{0_1} \Omega_\xi, (F_1 \times F_2)) &= J_2(\pi_{0_1} \Omega_\xi, F_2) = 0 \\ \text{et } J(\emptyset, (F_1 \times F_2)) &= J_2(\emptyset, F_2) = +\infty \end{aligned}$$

Nous avons évidemment les propriétés suivantes :

II.3.2.- PROPOSITION.

Les informations J_2 et J de la proposition II.3.1. vérifient les relations suivantes : pour tout $(C, F_1, F_2) \in \pi_{0_1} S_\xi \times F_1 \times F_2$ et $F_1 \neq \emptyset$

$$J_2(C_{F_1}, F_2) = J(C, (F_1 \times F_2)) = J(C_{F_1}, (0_1 \times F_2))$$

$$J_2(C, F_2) = J(C, (0_1 \times F_2))$$

Inversement, soit une information J sur $(\prod_{0_1} \Omega_\xi \times (0_1 \times 0_2), \otimes_{0_1} S_\xi \times (F_1 \times F_2))$ au sens I.2.2. - I.2.6. avec F_{12}^0 la classe des groupes d'observateurs de $F_1 \times F_2$ qui n'apportent pas d'information pour J (nous n'imposons a priori aucune forme particulière à F_{12}^0) ; définissons une information J_2 sur $(\prod_{0_1} \Omega_\xi \times 0_2, \otimes_{0_1} S_\xi \times F_2)$ au sens I.2.2. - I.2.6. en posant pour tout $(C, F_2) \in \otimes_{0_1} S_\xi \times F_2$.

$$J_2(C, F_2) = J(C, (0_1 \times F_2)).$$

La classe $F_2^0 = \{F_2 \in F_2 \mid (0_1, F_2) \in F_{12}^0\}$ qui vérifie les hypothèses de I.2.6. est la classe des groupes d'observateurs de F_2 qui n'apportent pas d'information pour J_2 . En règle générale, l'information J' sur $(\prod_{0_1} \Omega_\xi \times (0_1 \times 0_2), \otimes_{0_1} S_\xi \times (F_1 \times F_2))$ construite à partir de J_2 suivant la méthode de la proposition II.3.1. ne coïncide pas avec l'information J donnée ; nous avons cependant le résultat suivant :

II.3.3.- PROPOSITION.

Avec les notations ainsi définies une condition nécessaire et suffisante pour que les informations J et J_2 vérifient les relations de la proposition

II.3.1. est que, pour tout $(C, F_1, F_2) \in \otimes_{0_1} S_\xi \times F_1 \times F_2$ et $F_1 \neq \emptyset$:

$$J(C, (F_1 \times F_2)) = J(C_{F_1}, (0_1 \times F_2)).$$

La condition est nécessaire par suite de la proposition II.3.2. et il est facile de montrer qu'elle est aussi suffisante. Notons que la condition II.3.3. implique en particulier pour tout $(F_1, F_2) \in F_1 \times F_2$ et $F_1 \neq \emptyset$

$$J(\emptyset, (F_1 \times F_2)) = J(\emptyset, (0_1 \times F_2)) = J_2(\emptyset, F_2) = +\infty \text{ si } F_2 \notin F_2^0$$

$$J(\emptyset, (F_1 \times F_2)) = J(\emptyset, (0_1 \times F_2)) = J_2(\emptyset, F_2) = 0 \text{ si } F_2 \in F_2^0$$

ce qui entraîne effectivement que la classe F_{12}^0 sur laquelle nous n'avons

fait aucune hypothèse doit être de la forme désirée :

$$F_{12}^0 = \{(F_1, F_2) \mid F_1 = \emptyset \text{ ou } F_2 \in F_2^0\}$$

REMARQUES SUR LES NOTIONS DE DEPENDANCE ET D'INDEPENDANCE.

Pour étudier les relations entre les notions d'indépendance et de dépendance pour les informations J et J_2 , nous allons supposer que la condition de la proposition II.3.3. est vérifiée ; c'est-à-dire que nous avons construit l'information J sur $(\prod_{\emptyset} \Omega_{\xi} \times (O_1 \times O_2), \bigotimes_{\emptyset} S_{\xi} \times (F_1 \times F_2))$ à partir de l'information J_2 sur $(\prod_{\emptyset} \Omega_{\xi} \times O_2, \bigotimes_{\emptyset} S_{\xi} \times F_2)$ suivant la technique indiquée dans la proposition II.3.1.

Soient deux éléments (C, F_1, F_2) et (C', F'_1, F'_2) de $\bigotimes_{\emptyset} S_{\xi} \times F_1 \times F_2$; suivant la définition I.2.3., ils sont J -indépendants en information si

$$J[(C \cap C'), ((F_1 \times F_2) \cup (F'_1 \times F'_2))] = J(C, (F_1 \times F_2)) + J(C', (F'_1 \times F'_2))$$

En général la réunion de $(F_1 \times F_2)$ et $(F'_1 \times F'_2)$ n'est pas un pavé mesurable et dans ce cas le premier membre de cette relation n'a pas de sens. Nous devons donc nous limiter aux éléments tels que $(F_1 \times F_2) \cup (F'_1 \times F'_2)$ est un pavé mesurable.

Etant donné $F_2 = F'_2 \in F_2 - F_2^0$, il est facile de voir, comme dans le paragraphe II.1., que la J -indépendance des éléments $(C, (F_1 \times F_2))$ et $(C', (F'_1 \times F_2))$ n'implique pas en général la J_2 -indépendance des éléments (C_{F_1}, F_2) et $((C')_{F'_1}, F_2)$ et inversement. Cependant, on a le résultat suivant : soient F_1, F'_1 deux groupes d'observateurs donnés de F_1 ; pour tout $C \in \bigotimes_{F_1} S_{\xi}$ et $C' \in \bigotimes_{F'_1} S_{\xi}$ la J -indépendance des éléments $(C, (F_1 \times F_2))$ et $(C', (F'_1 \times F_2))$ implique la J_2 -indépendance des éléments (C_{F_1}, F_2) et $((C')_{F'_1}, F_2)$ et inversement.

Suivant le même type de raisonnement, on pourrait généraliser facilement les autres résultats du paragraphe II.1., c'est-à-dire préciser les liens entre la J - σ -indépendance et la J_2 - σ -indépendance d'une part, et les relations entre les informations conditionnelles définies à partir de J et de J_2 d'autre part.

II.4.- GENERALISATION.

Reprenons les notations du paragraphe I.1.3., pour simplifier les énoncés, nous avons supposé que les espaces mesurables (Ω, S) et $(\hat{\Omega}, \hat{S})$ sont formés par les σ -algèbres de Boole S et $\hat{S}$ de parties de Ω et $\hat{\Omega}$. Dans la mesure où l'on n'utilise pas la loi de composition (I.1.1.) il n'est pas utile d'imposer aux ensembles de parties S et $\hat{S}$ d'être fermés pour la réunion finie ou dénombrable. De même, moyennant certaines précautions dans la définition et l'utilisation des notions de dépendance et d'indépendance, il n'est pas nécessaire de supposer que les complémentaires de tous les ensembles de S (resp. $\hat{S}$) appartiennent à S (resp. $\hat{S}$). Il est clair que *tous les énoncés du chapitre I restent vrais quand on suppose simplement que S est un ensemble de parties de Ω contenant les ensembles $\emptyset, \Omega$ et fermé pour l'intersection finie.*

L'intersection des éléments de S est utilisée dans les définitions de l'information conditionnelle et des notions d'indépendance. *On exigera que S soit fermé pour l'intersection dénombrable seulement si l'on utilise les notions de σ -indépendance.*

Dans ce deuxième chapitre, nous avons supposé que $(\hat{\Omega}, \hat{S})$ est l'espace produit des espaces probabilisables $\{(\Omega_\xi, S_\xi) \mid \xi \in O\}$ comme on le définit classiquement en probabilité ; $\hat{S}$ est alors la σ -algèbre de Boole de parties de $\hat{\Omega}$ engendrée par les pavés mesurables, c'est-à-dire par les éléments de la forme $\prod_{\xi \in O} C_\xi$, où les parties $C_\xi \in S_\xi$ de Ω_ξ sont supposées n'être

distinctes de Ω_ξ que pour un nombre fini de $\xi \in O$. On peut aussi définir $\hat{S}$ à partir du produit des $\prod_{\xi \in O} S_\xi$; c'est-à-dire comme la famille de parties de $\hat{\Omega}$ composée des éléments de la forme $\prod_{\xi \in O} C_\xi$, $C_\xi \in S_\xi$, les C_ξ pouvant être tous, ou en partie, distincts de Ω_ξ .

Tous les énoncés du chapitre II restent vrais quand on suppose que, pour tout $\xi \in O$, S_ξ est un ensemble de parties de Ω_ξ contenant $\emptyset$ et Ω_ξ et fermé pour l'intersection finie avec : $\hat{S} = \{ \prod_{\xi \in O} C_\xi \mid C_\xi \in S_\xi, \xi \in O \}$

Effectivement, il est clair que $\emptyset$ et $\hat{\Omega}$ appartiennent à $\hat{S}$ et que $\hat{S}$ est fermé pour l'intersection finie ; dans ce cas, pour tout couple $(\hat{C}, F) \in \hat{S} \times F$, le cylindre $\hat{C}_F$ appartient à $\hat{S}$. On exigera, le cas échéant, la fermeture pour l'intersection dénombrable des ensembles S_ξ de parties de Ω_ξ , $\xi \in O$, si l'on utilise la notion de σ -indépendance.

Terminons par une propriété, très importante dans la pratique, de l'information I au sens de J. KAMPE DE FERIET. Supposons que l'ensemble Ω est fini ; soit A_1 un ensemble fini de parties de Ω contenant Ω , $\emptyset$ et fermé pour l'intersection et A_2 un autre ensemble fini de parties tel que $A_1 \subseteq A_2$. Etant donné une information I_2 sur (Ω, A_2) il est trivial de définir sur (Ω, A_1) l'information I_1 restriction de I_2 à A_1 . Inversement étant donné I_1 sur (Ω, A_1) que peut-on dire des informations I_2 sur (Ω, A_2) dont I_1 est la restriction sur (Ω, A_1) ? Soit $A_2 \in A_2$:

- Si $A_2 \in A_1$ alors $I_2(A_2) = I_1(A_2)$
- Si $A_2 \notin A_1$ définissons

$$A'_2 = \{ A_1 \mid A_1 \in A_1, A_1 \supseteq A_2 \} \in A_1$$

On a

$$I_2(A_2) \geq I_1(A'_2) .$$

En particulier si l'on pose $I'_1(A_2) = I_1(A'_2)$ pour tout $A_2 \in A_2$, il est facile de montrer que I'_1 est une information au sens de I.1.1. sur (Ω, A_2) , qui est égale à I_1 sur (Ω, A_1) et que toute information I_2 sur (Ω, A_2) qui coïncide avec I_1 sur (Ω, A_1) vérifiera

$$A_2 \in A_2 \quad I_2(A_2) \geq I'_1(A_2)$$

En quelque sorte I'_1 définit la "plus petite" information possible sur (Ω, A_2) qui prolonge I_1 sur (Ω, A_1) .

C'est la propriété que nous utilisons au paragraphe IV.3.12 pour définir l'information I'' , ainsi qu'au paragraphe VI.4.1. Si l'on connaît une information I définie sur l'ensemble des pavés mesurables de $\hat{\Omega}$ ($\hat{\Omega}$ fini), on peut prolonger I sur la tribu engendrée $\hat{S}$ en définissant l'information d'un élément $\hat{C} \in \hat{S}$ comme l'information apportée par le plus petit pavé mesurable qui contient $\hat{C}$. De même dans le paragraphe II.2, nous avons défini I sur A_T et non sur S tout entier, on peut le cas échéant prolonger I sur S en définissant l'information apportée par un élément $A \in S$ comme l'information apportée par le plus petit élément de A_T , s'il existe, qui contient A .

CHAPITRE III

DEPENDANCE ET INDEPENDANCE
DES OBSERVATEURS.

Nous étudions d'abord, dans ce chapitre, la forme générale de l'information J sur $(\Omega \times O, S \times F)$ quand on suppose que l'information apportée par deux groupes disjoints d'observateurs du même événement est la somme des informations apportées par chacun de ces groupes. Nous distinguerons le cas où l'ensemble O des observateurs est fini et celui où il est quelconque. Sous cette hypothèse d'indépendance, l'information $J(A, F)$ apportée sur l'événement $A \in S$ par le groupe $F \in F$ d'observateurs est égale à la moyenne des informations apportées sur A par chaque observateur $\xi \in F$, ce qui implique en particulier que l'information mutuelle entre les observateurs ξ de F est nulle.

Cette première hypothèse nous paraît assez restrictive ; aussi proposons-nous, dans la deuxième partie de ce chapitre, une méthode plus générale de construction de l'information : nous imaginons qu'il existe une partition (O_1, O_2) de l'ensemble O , les observateurs de O_1 (respectivement O_2) étant indépendants entre eux. On montre alors que l'information $J(A, (F_1 \cup F_2))$ apportée sur un événement $A \in S$ par le groupe $F_1 \cup F_2$ ($F_1 \subset O_1, F_2 \subset O_2$) d'observateurs est fonction des informations $J(A, \{\xi\})$ et $J(A, \{\xi_1, \xi_2\})$ apportées sur A par chaque observateur $\xi \in F_1 \cup F_2$ et chaque couple d'observateurs $(\xi_1, \xi_2) \in F_1 \times F_2$.

Les méthodes de construction des informations I et J ainsi déterminées sont à la base des applications envisagées dans la suite de ce travail. Ces résultats nous permettront aussi de mieux situer nos travaux par rapport à ceux de J. KAMPE DE FERIET sur "Les mesures de l'information par un ensemble d'observateurs indépendants", [14], [15], [17], ceux de B. FORTE sur les "Collecteurs" [8] et aux résultats de B. SCHWEIZER et A. SKLAR sur "Les mesures aléatoires d'information" [30], [31].

III.1. - CAS D'UN ENSEMBLE FINI D'OBSERVATEURS INDEPENDANTS.

Reprenons les notations du paragraphe I.2 : soient Ω un ensemble quelconque, non vide, et S une σ -algèbre de Boole de parties de Ω ; notons O l'ensemble supposé fini des observateurs, $F = P(O)$ l'algèbre de Boole des parties de O . Soit une information J au sens I.2.2. et I.2.6. sur $(\Omega \times O, S \times F)$, pour laquelle tous les observateurs $\xi \in O$ sont supposés non négligeables, c'est-à-dire d'après la définition I.2.6. et par suite de la proposition I.2.7. telle que $F_O = \{\emptyset\}$.

Pour tout $\xi \in O$ fixé, notons I_ξ l'application de S dans $\bar{\mathbb{R}}^+$ définie pour tout $A \in S$ par

$$I_\xi(A) = J(A, \{\xi\}).$$

Les définitions I.2.2. et I.2.6. impliquent que I est une information au sens I.1.1. sur (Ω, S) .

Supposons que tous les groupes disjoints d'observateurs de F sont J -indépendants en information au sens de la définition I.2.3., c'est-à-dire que pour tout couple $(F_1, F_2) \in F \times F$ et $F_1 \cap F_2 = \emptyset$ nous avons :

$$\forall A \in S \Rightarrow J(A, \{F_1 \cup F_2\}) = J(A, F_1) + J(A, F_2).$$

De proche en proche et en un nombre fini d'étapes, nous obtenons pour tout $(A, F) \in S \times F$

$$J(A, F) = \sum_{\xi \in F} J(A, \{\xi\}) = \sum_{\xi \in F} I_{\xi}(A).$$

Inversement, il est clair que si cette relation est vraie pour tout $(A, F) \in S \times F$ tous les groupes disjoints d'observateurs sont J -indépendants.

Dans le cas fini, il n'est pas nécessaire de supposer qu'il existe un système de pondération sur les observateurs. Cependant, pour que les énoncés suivants soient comparables à ceux du paragraphe III.3, définissons une mesure positive finie λ sur (Ω, F) , de masse totale égale à 1, telle que pour tout $\xi \in \Omega$, $\lambda(\xi) > 0$. On sait que si l'on multiplie une information au sens de J. KAMPE DE FERIET par une constante donnée strictement positive, on obtient une nouvelle information, ce qui nous permet d'énoncer la proposition suivante.

III.1.1.- PROPOSITION.

Etant donné une information J sur $(\Omega \times \Omega, S \times F)$ avec $F_0 = \{\emptyset\}$, une condition nécessaire et suffisante pour que tous les groupes disjoints d'observateurs de F soient J -indépendants en information est que pour tout $(A, F) \in S \times F$; on ait : $J(A, F) = \sum_{\xi \in F} I_{\xi}(A)$

$$(\text{respectivement : } J(A, F) = \sum_{\xi \in F} \lambda(\xi) \cdot I'_{\xi}(A))$$

où I_ξ (resp. I'_ξ) est l'information au sens I.1.1. sur (Ω, S) définie par

$$I_\xi(.) = J(. , \{\xi\})$$

$$(\text{resp. } I'_\xi(.) = \frac{1}{\lambda(\xi)} \cdot J(. , \{\xi\})).$$

De plus, il est clair que

III.1.2.- COROLLAIRE.

Soit $\{I_\xi | \xi \in O\}$ une famille finie d'informations quelconques au sens I.1.1. sur (Ω, S) . L'application J définie par :

$$\forall (A, F) \in S \times F, \quad J(A, F) = \sum_{\xi \in F} \lambda(\xi) \cdot I_\xi(A)$$

est une information sur $(\Omega \times O, S \times F)$, avec $F_0 = \{\emptyset\}$, telle que tous les groupes disjoints d'observateurs sont J -indépendants.

REMARQUE : CAS OU $F_0 \neq \{\emptyset\}$.

Si F_0 est différent de $\{\emptyset\}$ pour l'information J sur $(\Omega \times O, S \times F)$ et si tous les groupes disjoints d'observateurs sont J -indé-

pendants, on obtient encore pour tout $(A, F) \in S \times F$, la relation

$$J(A, F) = \sum_{\xi \in F} J(A, \{\xi\}).$$

Partitionnons l'ensemble O en deux groupes : les observateurs J -négligeables (soit O_0) et les autres (soit O_1) :

$$O_0 = \{\xi \in O \mid \{\xi\} \in F_0\}$$

$$O_1 = \{\xi \in O \mid \{\xi\} \notin F_0\}$$

Il est clair que $F_0 = P(O_0)$ (ce qui pourrait ne pas être vrai si les groupes disjoints d'observateurs n'étaient pas J -indépendants) et la proposition III.1.1. reste vraie en prenant les sommations sur $F \cap O_1$.

Inversement, comme dans le corollaire III.1.2., étant donné une mesure de pondération λ quelconque sur (O, F) , définissons O_0 comme l'ensemble des observateurs de O λ -négligeables :

$$O_0 = \{\xi \in O \mid \lambda(\xi) = 0\}$$

et posons que $\lambda(\xi) = 0$ implique $\lambda(\xi) \cdot I_\xi(A) = 0$ pour tout $A \in S$; alors

l'application J définie sur $(\Omega \times O, S \times F)$ par $J(A, F) = \sum_{\xi \in F} \lambda(\xi) \cdot I_\xi(A)$

est une information ; $F_0 = P(O_0)$, et tous les groupes disjoints d'observateurs sont J -indépendants.

INFORMATIONS SUR $(\hat{\Omega} \times O, \hat{S} \times F)$.

Reprenons les notations du paragraphe II.1 et supposons que l'information J sur $(\hat{\Omega} \times O, \hat{S} \times F)$ vérifie la condition du théorème II.1.4. Notons

I l'information au sens de J. KAMPE DE FERIET sur $(\hat{\Omega}, \hat{S})$ définie par

$$\forall \hat{C} \in \hat{S}, \quad I(\hat{C}) = J(\hat{C}, \emptyset).$$

Nous avons par hypothèse pour tout $(\hat{C}, F) \in \hat{S} \times F$:

$$\begin{aligned} J(\hat{C}, F) &= J(\hat{C}_F, \emptyset) = I(\hat{C}_F) & \text{si } F \neq \emptyset \\ J(\hat{C}, F) &= 0 & \text{si } F = \emptyset, \end{aligned}$$

Soit $\hat{C}$ un élément de $\hat{S}$, nous notons $\{C_{\{\xi\}} | \xi \in \emptyset\}$ les projections supposées S_{ξ} -mesurables, de $\hat{C}$ sur Ω_{ξ} et $\{\hat{C}_{\{\xi\}} | \xi \in \emptyset\}$ les cylindres associés (I.1.3.). Avec les notations ainsi définies, il vient :

III.1.3.- PROPOSITION.

Une condition nécessaire et suffisante pour que tous les groupes disjoints d'observateurs de F soient J -indépendants en information est que pour tout $(\hat{C}, F) \in \hat{S} \times F$

$$J(\hat{C}, F) = \sum_{\xi \in F} I(\hat{C}_{\{\xi\}}).$$

La démonstration est évidente en remarquant que

$J(\hat{C}, \{\xi\}) = I(\hat{C}_{\{\xi\}})$ et en appliquant la proposition III.1.1. Comme

$\prod_{\xi \in \emptyset} C_{\{\xi\}} = \bigcap_{\xi \in \emptyset} \hat{C}_{\{\xi\}}$, il vient :

III.1.4.- COROLLAIRE.

Si la condition de la proposition III.1.3. est vérifiée, pour tout $(\hat{C}, F) \in \hat{S} \times F$ nous avons $I(\hat{C}) = I(\prod_{\xi \in \emptyset} C_{\{\xi\}}) = I(\bigcap_{\xi \in \emptyset} \hat{C}_{\{\xi\}})$

et $J(\hat{C}, F) = J(\bigcap_{\xi \in \emptyset} \hat{C}_{\{\xi\}}, F) = I(\bigcap_{\xi \in F} \hat{C}_{\{\xi\}}).$

Inversement soient $\{(\Omega_\xi, \mathcal{S}_\xi, I_\xi) \mid \xi \in \mathcal{O}\}$ une famille finie d'espaces mesurables $(\Omega_\xi, \mathcal{S}_\xi)$ sur lesquels on définit une information I_ξ au sens I.1.1. et une mesure positive λ_ξ de masse 1, sur $(\mathcal{O}, \mathcal{F})$ telle que, pour tout $\xi \in \mathcal{O}$, $\lambda(\xi) > 0$.

III.1.5.- PROPOSITION.

L'application J définie par

$$\forall F \in \mathcal{F}, \forall (C_\xi)_{\xi \in \mathcal{O}} \in \prod_{\xi \in \mathcal{O}} \mathcal{S}_\xi, \quad J(\prod_{\xi \in \mathcal{O}} C_\xi, F) = \sum_{\xi \in \mathcal{F}} \lambda(\xi) \cdot I_\xi(C_\xi)$$

est une information sur les pavés mesurables de $(\hat{\Omega}, \hat{\mathcal{S}})$ vérifiant $F_0 = \{\emptyset\}$. Cette information peut se prolonger à $(\hat{\Omega} \times \mathcal{O}, \hat{\mathcal{S}} \times \mathcal{F})$ en posant pour tout $(\hat{C}, F) \in \hat{\mathcal{S}} \times \mathcal{F}$:

$$J(\hat{C}, F) = J(\prod_{\xi \in \mathcal{O}} C_{\{\xi\}}, F) .$$

Elle est telle que tous les groupes disjoints d'observateurs sont J -indépendants.

III.1.6.- COROLLAIRE.

L'information J sur $(\hat{\Omega} \times \mathcal{O}, \hat{\mathcal{S}} \times \mathcal{F})$ définie dans la proposition III.1.5. vérifie les hypothèses du théorème II.1.4. ; en particulier, elle permet de définir sur $(\hat{\Omega}, \hat{\mathcal{S}})$ une information I en posant :

$$\forall \hat{C} \in \hat{\mathcal{S}}, \quad I(\hat{C}) = J(\hat{C}, \mathcal{O})$$

et l'on a pour tout $(\hat{C}, F) \in \hat{S} \times F$ et tout $\xi \in O$:

$$J(\hat{C}, F) = J(\hat{C}_F, O) = I(\hat{C}_F).$$

$$J(\hat{C}, \{\xi\}) = J(\hat{C}_{\{\xi\}}, O) = I(\hat{C}_{\{\xi\}}) = \lambda(\xi) \cdot I_{\xi}(C_{\{\xi\}}).$$

III.2. - AUTRES HYPOTHESES D'INDEPENDANCE.

Le principal résultat du paragraphe III.1 est celui-ci :
si l'on suppose que tous les groupes disjoints d'observateurs sont indépendants,
il suffit de connaître les informations $J(A, \{\xi\})$ apportées par chaque
observateur $\xi \in O$ sur l'événement $A \in S$ pour pouvoir calculer l'information
apportée sur cet événement par un groupe F quelconque d'observateurs.

$$J(A, F) = \sum_{\xi \in F} J(A, \{\xi\}).$$

En particulier, pour tout couple (ξ_1, ξ_2) d'observateurs de O , on a :

$$J(A, \{\xi_1, \xi_2\}) = J(A, \{\xi_1\}) + J(A, \{\xi_2\})$$

ou encore

$$J((A, \{\xi_1\}) / (A, \{\xi_2\})) = J(A, \{\xi_1\}).$$

La nouvelle hypothèse, que nous proposons, a été conçue en fonction des objectifs suivants :

a) Il peut exister des couples (ξ_1, ξ_2) d'observateurs pour lesquels

$$J(A, \{\xi_1, \xi_2\}) \neq J(A, \{\xi_1\}) + J(A, \{\xi_2\})$$

$$\text{ou } J[(A, \{\xi_1\})/(A, \{\xi_2\})] \neq J(A, \{\xi_1\}).$$

b) L'information $J(A, F)$ peut être calculée en connaissant les informations $J(A \times \{\xi\})$ et $J(A \times \{\xi_1, \xi_2\})$.

c) L'hypothèse d'indépendance du paragraphe III.1 doit être un cas particulier de cette hypothèse plus générale.

Nous cherchons donc à déterminer un modèle moins restrictif que celui du paragraphe III.1 mais qui aboutisse lui aussi à des formules suffisamment simples pour être utilisables. Nous proposerons, dans la partie V.3., d'autres directions de travail possibles et des éléments de discussion de ces différentes hypothèses.

Soient (Ω, S) l'espace quelconque des événements, O l'ensemble supposé fini des observateurs, $F = P(O)$ et J une information sur $(\Omega \times O, S \times F)$ telle que $F_0 = \{\emptyset\}$, vérifiant de plus :

$$\forall A \in S, A \neq \emptyset \Rightarrow J(A, O) < +\infty.$$

Etant donnée une partition (O_1, O_2) de O , nous notons : $F_i = P(O_i)$, $i = 1, 2$.

Dans l'hypothèse dont l'énoncé suit, nous supposons que les groupes disjoints d'observateurs de O_i sont indépendants pour l'observation d'un événement quelconque $A \in S$ quelle que soit l'information apportée sur cet

événement par les observateurs de O_2 , et inversement.

III.2.1.- HYPOTHESE.

Pour tout $(F_1, F'_1) \in F_1 \times F_1$ tel que $F_1 \cap F'_1 = \emptyset$; pour tout $F_2 \in F_2$ et tout $A \in S$, $A \neq \emptyset$; nous avons

$$J([A, (F_1 \cup F'_1)]/[A, F_2]) = J([A, F_1]/[A, F_2]) + J([A, F'_1]/[A, F_2])$$

de même en permutant les rôles de (O_1, F_1) et (O_2, F_2) .

Les expressions écrites ont toujours un sens car nous avons supposé que pour tout événement $A \in S$ et $A \neq \emptyset$

$$J(A, \emptyset) < +\infty$$

donc en particulier

$$J(A, F_2) < +\infty.$$

En prenant $F_2 = \emptyset$, il vient :

$$J(A, (F_1 \cup F'_1)) = J(A, F_1) + J(A, F'_1).$$

Cette hypothèse implique donc que les groupes disjoints d'observateurs de O_1 (respectivement de O_2) sont indépendants en information au sens du paragraphe III.1. Nous avons donc pour $F_i \in F_i$, $i = 1, 2$:

$$J(A, F_i) = \sum_{\xi \in F_i} J(A, \{\xi\}) \quad i = 1, 2.$$

Conformément à la définition I.3.5., nous appelons *information mutuelle* apportée sur le même événement A par les groupes F et F' d'observateurs de $\mathcal{O}$, la quantité

$$J((A, F); (A, F')) = J(A, F) + J(A, F') - J(A, (F \cup F'))$$

$$\text{si } J(A, F) < +\infty \text{ et } J(A, F') < +\infty.$$

Dans ce paragraphe, l'information mutuelle est parfaitement définie pour tout $A \in S$, $A \neq \emptyset$.

III.2.2.- PROPOSITION.

L'hypothèse III.2.1. est vérifiée si et seulement si pour tout $(F_1^1, F_1^2) \in F_1 \times F_1$, $F_1^1 \cap F_1^2 = \emptyset$, $i = 1, 2$, et pour tout : $A \in S$, $A \neq \emptyset$, on a :

$$J(A, (\bigcup_{\substack{i=1,2 \\ j=1,2}} F_1^j)) = \sum_{\substack{i=1,2 \\ j=1,2}} J(A, F_1^j)$$

$$- \sum_{\substack{k=1,2 \\ j=1,2}} J((A, F_1^j); (A, F_2^k))$$

La condition est suffisante ; en effet, les quantités de la forme $J((A, F); (A, \emptyset))$ étant nulles, il est évident que la relation de la proposition III.2.2. implique

$$J(A, (F_1^1 \cup F_1^2)) = J(A, F_1^1) + J(A, F_1^2) \text{ et}$$

$$J(A, (F_2^1 \cup F_2^2)) = J(A, F_2^1) + J(A, F_2^2).$$

D'autre part en posant : $F_2 = F_2^1$ et $F_2^2 = \emptyset$, il vient :

$$J(A, (F_1^1 \cup F_1^2)) / (A, F_2) = J(A, (F_1^1 \cup F_1^2 \cup F_2)) - J(A, F_2)$$

$$= J(A, F_1^1) - J((A, F_1^1); (A, F_2)) + J(A, F_1^2) - J((A, F_1^2); (A, F_2)),$$

c'est-à-dire :

$$J((A, (F_1^1 \cup F_1^2)) / (A, F_2)) = J((A, F_1^1) / (A, F_2)) + J((A, F_1^2) / (A, F_2))$$

de même en permutant les rôles de (O_1, F_1) et (O_2, F_2) .

La condition est nécessaire ; en effet, avec les notations de la proposition III.2.2. et en posant : $F_2 = F_2^1 \cup F_2^2$, la relation de l'hypothèse III.2.1. s'écrit

$$J(A, (F_1^1 \cup F_1^2 \cup F_2)) = J(A, F_2) + \sum_{j=1,2} J(A, F_1^j) - \sum_{j=1,2} J((A, F_1^j); (A, F_2))$$

ajoutons et retranchons $J(A, F_2)$

$$J(A, (F_1^1 \cup F_1^2 \cup F_2)) = \sum_{j=1,2} J(A, F_1^j) - J(A, F_2) + \sum_{j=1,2} J((A, F_2) / (A, F_1^j)).$$

On sait par hypothèse que pour $j = 1, 2$:

$$\begin{aligned}
J((A, F_2)/(A, F_1^j)) &= J((A, F_2^1)/(A, F_1^j)) + J((A, F_2^2)/(A, F_1^j)) \\
&= \sum_{k=1,2} J(A, F_2^k) - \sum_{k=1,2} J((A, F_1^j); (A, F_2^k))
\end{aligned}$$

comme nous avons montré que l'hypothèse III.2.1. implique aussi :

$$J(A, F_2) = \sum_{k=1,2} J(A, F_2^k)$$

on obtient la relation désirée en reportant ces valeurs dans l'expression précédente.

III.2.3.- COROLLAIRE.

Avec les notations de la proposition précédente en posant

$F_1 = F_1^1 \cup F_1^2$ ou $F_2 = F_2^1 \cup F_2^2$, l'hypothèse III.2.1. implique

$$J((A, (F_1^1 \cup F_1^2)); (A, F_2)) = J((A, F_1^1); (A, F_2)) + J((A, F_1^2); (A, F_2))$$

$$J((A, F_1); (A, (F_2^1 \cup F_2^2))) = J((A, F_1); (A, F_2^1)) + J((A, F_1); (A, F_2^2))$$

$$J((A, (F_1^1 \cup F_1^2)); (A, (F_2^1 \cup F_2^2))) = \sum_{\substack{k=1,2 \\ j=1,2}} J((A, F_1^j); (A, F_2^k)).$$

La démonstration est évidente à partir de la proposition III.2.2.

Notons que ces relations pourraient être vraies sans que

$$J(A, (F_1^1 \cup F_1^2)) = J(A, F_1^1) + J(A, F_1^2)$$

c'est-à-dire sans que l'hypothèse III.2.1. soit réalisée.

III.2.4.- THEOREME.

Une condition nécessaire et suffisante pour que l'hypothèse III.2.1. soit vérifiée est que pour tout $F_i \in \mathcal{F}_i$, $i = 1, 2$, et pour tout $A \in S$, $A \neq \emptyset$, on ait :

$$J(A, (F_1 \cup F_2)) = \sum_{\xi \in F_1} J(A, \{\xi\}) + \sum_{\xi \in F_2} J(A, \{\xi\}) - \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} J((A, \{\xi_1\}) ; (A, \{\xi_2\})).$$

La condition est évidemment suffisante car elle implique la condition de la proposition III.2.2. On montre qu'elle est aussi nécessaire par une démonstration analogue à celle de la condition nécessaire de la proposition III.2.2. Ce théorème peut aussi s'énoncer de la manière suivante :

III.2.5.- COROLLAIRE.

L'hypothèse III.2.1. est vérifiée si et seulement si pour tout $A \in S$, $A \neq \emptyset$, et pour tout $F_i \in \mathcal{F}_i$, $i = 1, 2$, on a :

$$J(A, F_i) = \sum_{\xi \in F_i} J(A, \{\xi\}) \quad i = 1, 2$$

$$J((A, F_1) ; (A, F_2)) = \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} J((A, \{\xi_1\}) ; (A, \{\xi_2\})).$$

Citons également le résultat suivant :

III.2.6.- COROLLAIRE.

Etant donné une information J sur $(\Omega \times O, S \times F)$ et une partition (O_1, O_2) de O , une condition nécessaire et suffisante pour que tous les groupes disjoints d'observateurs soient indépendants (au sens du paragraphe III.1) est que J vérifie les hypothèses de III.2.1. et que tous les couples $(\xi_1, \xi_2) \in O_1 \times O_2$ d'observateurs soient indépendants.

EXEMPLE.

Nous allons maintenant donner une méthode de construction d'une information vérifiant l'hypothèse III.2.1.

Etant donnée une partition (O_1, O_2) de O , notons $F_i = P(O_i)$, $i = 1, 2$, attachons à chaque observateur $\xi \in O$ et à chaque couple d'observateurs $(\xi_1, \xi_2) \in O_1 \times O_2$ des informations sur (Ω, S) notées respectivement I_ξ et $I_{\xi_1 \xi_2}$. Supposons que la famille d'informations ainsi définie vérifie pour tout $(\xi_1, \xi_2) \in O_1 \times O_2$ et pour tout $A \in S$, $A \neq \emptyset$:

$$\sup(I_{\xi_1}(A), I_{\xi_2}(A)) \leq I_{\xi_1 \xi_2}(A) < +\infty.$$

Soit λ une mesure positive de masse totale égale à 1 sur $(O_1 \times O_2, F_1 \otimes F_2)$.

Nous écrirons $\lambda(\xi_1, \xi_2)$ la mesure du couple $(\xi_1, \xi_2) \in O_1 \times O_2$ et nous utiliserons les notations condensées suivantes pour $F_i \in F_1$, $\xi_i \in O_i$, $i = 1, 2$:

$$\lambda(F_1, F_2) = \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \quad \text{avec} \quad \lambda(F_1, \emptyset) = \lambda(\emptyset, F_2) = 0$$

$$\lambda_1(\xi_1) = \lambda(\xi_1, O_2) = \sum_{\xi_2 \in O_2} \lambda(\xi_1, \xi_2)$$

$$\lambda_2(\xi_2) = \lambda(0_1, \xi_2) = \sum_{\xi_1 \in 0_1} \lambda(\xi_1, \xi_2).$$

Supposons enfin que les mesures marginales sont strictement positives :

$$\xi_i \in 0_i \Rightarrow \lambda_i(\xi_i) > 0 \quad i = 1, 2.$$

III.2.7.- PROPOSITION.

Avec les notations ainsi précisées, l'application J définie ci-dessous est une information sur $(\Omega \times 0, S \times F)$ qui vérifie les hypothèses III.2.1.

$$\forall (A, F_1, F_2) \in S \times F_1 \times F_2, A \neq \emptyset$$

$$\begin{aligned} J(A, (F_1 \cup F_2)) &= \sum_{\xi_1 \in F_1} \lambda(\xi_1, 0_2 - F_2) \cdot I_{\xi_1}(A) \\ &+ \sum_{\xi_2 \in F_2} \lambda(0_1 - F_1, \xi_2) \cdot I_{\xi_2}(A) \\ &+ \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(A). \end{aligned}$$

Cette quantité s'écrit aussi :

$$\begin{aligned} J(A, (F_1 \cup F_2)) &= \sum_{\xi_1 \in F_1} \lambda_1(\xi_1) \cdot I_{\xi_1}(A) \\ &+ \sum_{\xi_2 \in F_2} \lambda_2(\xi_2) \cdot I_{\xi_2}(A) \\ &- \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1}(A) + I_{\xi_2}(A) - I_{\xi_1 \xi_2}(A)) \end{aligned}$$

On pose de plus

$$J(\emptyset, F) = +\infty \quad \text{si } F \neq \emptyset$$

$$J(\emptyset, F) = 0 \quad \text{si } F = \emptyset.$$

Les deux définitions sont évidemment identiques. On en déduit pour tout $\xi_i \in \mathcal{O}_i$, $i = 1, 2$ et pour tout $A \in S$,

$$J(A, \{\xi_i\}) = \lambda_i(\xi_i) \cdot I_{\xi_i}(A) \quad i = 1, 2$$

$$\begin{aligned} J(A, \{\xi_1, \xi_2\}) &= J(A, \{\xi_1\}) + J(A, \{\xi_2\}) \\ &\quad - \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1}(A) + I_{\xi_2}(A) - I_{\xi_1 \xi_2}(A)) \end{aligned}$$

c'est-à-dire

$$J((A, \{\xi_1\}) ; (A, \{\xi_2\})) = \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1}(A) + I_{\xi_2}(A) - I_{\xi_1 \xi_2}(A))$$

donc

$$\begin{aligned} J(A, (F_1 \cup F_2)) &= \sum_{\xi_1 \in F_1} J(A, \{\xi_1\}) + \sum_{\xi_2 \in F_2} J(A, \{\xi_2\}) \\ &\quad - \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} J((A, \{\xi_1\}) ; (A, \{\xi_2\})). \end{aligned}$$

La condition du théorème III.2.4. est donc vérifiée ; il nous reste à montrer que J est une information pour obtenir l'hypothèse III.2.1. Par suite des propriétés des informations au sens I.1.1., il est évident, par définition de J , que pour tout $(A_1, A_2) \in S \times S$, $A_1 \subseteq A_2$, et tout $F_i \in \mathcal{F}_i$, $i = 1, 2$, on a :

$$J(A_1, (F_1 \cup F_2)) \geq J(A_2, (F_1 \cup F_2)) \geq 0.$$

D'autre part, pour $A \in S$, $F_2 \in F_2$, $(F_1', F_1'') \in F_1 \times F_1$ et $F_1' \cap F_1'' = \emptyset$, notons $F_1 = F_1' \cup F_1''$ et montrons que

$$J(A, (F_1 \cup F_2)) \geq J(A, (F_1' \cup F_2)).$$

Exprimons la différence de ces deux termes

$$\begin{aligned} J(A, (F_1 \cup F_2)) - J(A, (F_1' \cup F_2)) &= \sum_{\xi_1 \in F_1''} \lambda_1(\xi_1) \cdot I_{\xi_1}(A) \\ &- \sum_{(\xi_1, \xi_2) \in F_1'' \times F_2} \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1}(A) + I_{\xi_2}(A) - I_{\xi_1 \xi_2}(A)) \\ &= \sum_{\xi_1 \in F_1''} \lambda(\xi_1, 0_2 - F_2) \cdot I_{\xi_1}(A) \\ &+ \sum_{(\xi_1, \xi_2) \in F_1'' \times F_2} \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1 \xi_2}(A) - I_{\xi_2}(A)). \end{aligned}$$

Cette différence est toujours positive car nous avons supposé :

$I_{\xi_1 \xi_2}(A) \geq I_{\xi_2}(A)$. A partir de ces résultats, il est facile de montrer que

J est une information au sens des définitions I.2.2. et I.2.6. avec :

$F_0 = \{\emptyset\}$. La proposition est ainsi démontrée. Nous avons de plus les résultats suivants :

III.2.8.- PROPOSITION.

Avec les notations de la proposition III.2.7., nous avons pour tout

$A \in S$, $A \neq \emptyset$, et pour tout $F_1 \in F_1$, $\xi_1 \in 0_1$, $i = 1, 2$:

$$J(A, 0) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(A)$$

$$J(A, F_i) = \sum_{\xi \in F_i} \lambda_i(\xi) \cdot I_{\xi}(A) \quad i = 1, 2$$

$$J(A, \{\xi_i\}) = \lambda_i(\xi_i) \cdot I_{\xi_i}(A) \quad i = 1, 2$$

$$J(A, \{\xi_1, \xi_2\}) = (\lambda_1(\xi_1) - \lambda(\xi_1, \xi_2)) \cdot I_{\xi_1}(A) + (\lambda_2(\xi_2) - \lambda(\xi_1, \xi_2)) \cdot I_{\xi_2}(A) + \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(A)$$

$$J((A, \{\xi_1\}) ; (A, \{\xi_2\})) = \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1}(A) + I_{\xi_2}(A) - I_{\xi_1 \xi_2}(A))$$

$$J(A, (F_1 \cup F_2)) = J(A, F_1) + J(A, F_2) -$$

$$\sum_{(\xi_1, \xi_2) \in F_1 \times F_2} J((A, \{\xi_1\}) ; (A, \{\xi_2\}))$$

$$J((A, F_2)/(A, O_1)) = \sum_{(\xi_1, \xi_2) \in O_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1 \xi_2}(A) - I_{\xi_1}(A)).$$

Remarquons une différence importante entre la proposition III.2.7. et le résultat comparable du corollaire III.1.2. et de la proposition III.1.1. Etant donné J vérifiant l'hypothèse III.2.1. et λ la mesure de pondération l'information J n'est pas nécessairement de la forme indiquée dans la proposition III.2.7. En effet, si l'on peut toujours définir les informations I_{ξ} en posant pour $\xi \in O_i$, $i = 1, 2$:

$$I_{\xi}(A) = \frac{1}{\lambda_i(\xi)} \cdot J(A, \{\xi\})$$

pour $\xi_i \in O_i$, $i = 1, 2$, les quantités

$$I_{\xi_1 \xi_2}(A) = \frac{1}{\lambda(\xi_1, \xi_2)} \{ J(A, \{\xi_1, \xi_2\}) - \frac{\lambda_1(\xi_1) - \lambda(\xi_1, \xi_2)}{\lambda_1(\xi_1)} \cdot J(A, \{\xi_1\}) - \frac{\lambda_2(\xi_2) - \lambda(\xi_1, \xi_2)}{\lambda_2(\xi_2)} \cdot J(A, \{\xi_2\}) \}$$

ne définissent pas nécessairement une information, par exemple si, comme cela est possible, $J(A, \{\xi_1, \xi_2\}) = J(A, \{\xi_1\}) = J(A, \{\xi_2\})$ et $\lambda_1(\xi_1) = \lambda_2(\xi_2)$, on peut avoir : $I_{\xi_1 \xi_2}(A) < 0$.

INFORMATIONS SUR $(\hat{\Omega} \times 0, \hat{S} \times F)$.

Reprenons les notations du paragraphe II.1 et supposons que l'information J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ vérifie la condition du théorème II.1.4. Nous avons donc une information I sur $(\hat{\Omega}, \hat{S})$ telle que pour tout $(\hat{C}, F) \in \hat{S} \times F$

$$I(\hat{C}) = J(\hat{C}, 0) \text{ et } J(\hat{C}, F) = J(\hat{C}_F, 0) = I(\hat{C}_F).$$

III.2.9.- PROPOSITION.

Etant donné, l'espace $(\hat{\Omega} \times 0, \hat{S} \times F, J)$ ainsi défini et $(0_1, 0_2)$ une partition de l'ensemble fini 0 , une condition nécessaire et suffisante pour que l'hypothèse III.2.1. soit vérifiée est que pour tout $F_i \in F_1$, $i = 1, 2$ et pour tout $\hat{C} \in \hat{S}$, $\hat{C} \neq \emptyset$, on ait :

$$\begin{aligned}
J(\hat{C}, (F_1 \cup F_2)) &= \sum_{\xi_1 \in F_1} I(\hat{C}_{\{\xi_1\}}) + \sum_{\xi_2 \in F_2} I(\hat{C}_{\{\xi_2\}}) - \\
&- \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} (I(\hat{C}_{\{\xi_1\}}) + I(\hat{C}_{\{\xi_2\}}) - I(\hat{C}_{\{\xi_1, \xi_2\}})).
\end{aligned}$$

La démonstration est évidente à partir du théorème III.2.4. en remarquant que $\forall \xi \in O, J(\hat{C}, \{\xi\}) = I(\hat{C}_{\{\xi\}})$

$$\forall (\xi_1, \xi_2) \in O_1 \times O_2, J(\hat{C}, \{\xi_1, \xi_2\}) = I(\hat{C}_{\{\xi_1, \xi_2\}}).$$

Contrairement au résultat du corollaire III.1.4. nous n'avons plus

$$J(\hat{C}, F) = J(\prod_{\xi \in O} (C_{\{\xi\}}), F)$$

car en règle générale $\hat{C}_{\{\xi_1, \xi_2\}} \neq \hat{C}_{\{\xi_1\}} \cap \hat{C}_{\{\xi_2\}}$.

Soient la partition (O_1, O_2) de l'ensemble fini O et les familles

$$\{(\Omega_\xi, S_\xi, I_\xi) \mid \xi \in O\}, \{(\Omega_{\xi_1} \times \Omega_{\xi_2}, S_{\xi_1} \otimes S_{\xi_2}, I_{\xi_1 \xi_2}) \mid (\xi_1, \xi_2) \in O_1 \times O_2\} \text{ dont}$$

chaque élément est un espace mesurable sur lequel on a défini une information vérifiant pour tout $\xi_i \in O_i, i = 1, 2, C_{\xi_1} \in S_{\xi_1}, C_{\xi_2} \in S_{\xi_2}$.

$$\begin{aligned}
I_{\xi_1 \xi_2}(C_{\xi_1} \times \Omega_{\xi_2}) &= I_{\xi_1}(C_{\xi_1}) \\
I_{\xi_1 \xi_2}(\Omega_{\xi_1} \times C_{\xi_2}) &= I_{\xi_2}(C_{\xi_2})
\end{aligned}$$

Soit λ la mesure de pondération sur $(O_1 \times O_2, F_1 \otimes F_2)$ précédemment définie (III.2.7.).

III.2.10.- PROPOSITION.

L'application J , définie ci-dessous, est une information sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ qui vérifie l'hypothèse III.2.1. et la condition du théorème II.1.4.

$$\forall (\hat{C}, F_1, F_2) \in \hat{S} \times F_1 \times F_2, \hat{C} \neq \emptyset$$

$$\begin{aligned} J(\hat{C}, (F_1 \cup F_2)) &= \sum_{\xi_1 \in F_1} \lambda(\xi_1, 0_2 - F_2) \cdot I_{\xi_1}(C_{\{\xi_1\}}) \\ &+ \sum_{\xi_2 \in F_2} \lambda(0_1 - F_1, \xi_2) \cdot I_{\xi_2}(C_{\{\xi_2\}}) \\ &+ \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(C_{\{\xi_1, \xi_2\}}) \end{aligned}$$

ce qui s'écrit aussi :

$$\begin{aligned} J(\hat{C}, (F_1 \cup F_2)) &= \sum_{\xi_1 \in F_1} \lambda_1(\xi_1) \cdot I_{\xi_1}(C_{\{\xi_1\}}) \\ &+ \sum_{\xi_2 \in F_2} \lambda_2(\xi_2) \cdot I_{\xi_2}(C_{\{\xi_2\}}) \\ &- \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot (I_{\xi_1}(C_{\{\xi_1\}}) + I_{\xi_2}(C_{\{\xi_2\}}) \\ &\quad - I_{\xi_1 \xi_2}(C_{\{\xi_1, \xi_2\}})) \end{aligned}$$

on pose de plus

$$J(\emptyset, F) = 0 \quad \text{si} \quad F = \emptyset$$

$$J(\emptyset, F) = +\infty \quad \text{si} \quad F \neq \emptyset$$

Ce résultat est une application immédiate de la proposition III.2.7.

Notons de plus les résultats suivants :

III.2.11. - PROPOSITION.

Avec les notations de la proposition III.2.10. nous avons pour tout $\hat{C} \in \hat{S}$, $\hat{C} \neq \emptyset$, et pour tout $F_i \in F_i$, $\xi_i \in O_i$, $i = 1, 2$:

$$I(\hat{C}) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2} [C_{\{\xi_1, \xi_2\}}]$$

et avec $F = F_1 \cup F_2$

$$J(\hat{C}, F) = I(\hat{C}_F) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2} [(C_F)_{\{\xi_1, \xi_2\}}]$$

III.3. - CAS D'UN ENSEMBLE QUELCONQUE D'OBSERVATEURS INDEPENDANTS.

Soit J une information, au sens I.2.2. et I.2.6., sur $(\Omega \times O, S \times F)$; Ω et O sont des ensembles quelconques non vides, S et F des σ -algèbres de Boole de parties de Ω et O . Pour généraliser les résultats du paragraphe III.1 à un ensemble quelconque (non nécessairement fini) d'observateurs O , nous supposons que les groupes disjoints d'observateurs de F sont σ -indépendants en information pour l'observation des événements de S ; c'est-à-dire, en appliquant la définition I.3.2., que pour toute famille finie ou dénombrable de groupes d'observateurs deux à deux disjoints $\{F_i \in F | i \in K\}$ et pour tout $A \in S$, on a :

$$J(A, (\cup_{i \in K} F_i)) = \prod_{i \in K} J(A, F_i).$$

Comme par définition $J(A, \emptyset) = 0$, cela revient à supposer que pour tout $A \in S$ fixé l'application $J(A, \cdot)$ est une mesure positive sur (O, F) .

Définissons sur (O, F) une mesure λ , positive, finie, de masse 1,

qui, pour tout $A \in S$, domine la mesure $J(A, \cdot)$, c'est-à-dire telle que :

$$N \in F \text{ et } \lambda(N) = 0 \implies J(A, N) = 0.$$

On note $J(A, \cdot) \ll \lambda$.

Soit N la classe des ensembles mesurables négligeables de (O, F, λ) ; il est clair que la condition $J(A, \cdot) \ll \lambda$ pour tout $A \in S$ est équivalente à $N \subseteq F_0$ où F_0 est la classe des groupes d'observateurs qui n'apportent pas d'information pour J (I.2.6.).

III.3.1.- PROPOSITION.

Etant donné un événement $A \in S$ fixé, une information J sur $(\Omega \times O, S \times F)$ et λ une mesure de probabilité sur (O, F) telles que $N \subseteq F_0$, une condition nécessaire et suffisante pour que tous les groupes disjoints d'observateurs de (O, F) soient σ -indépendants en information pour l'observation de $A \in S$ est qu'il existe une variable aléatoire réelle, notée $I(A, \cdot)$, unique, positive, définie à une λ -équivalence près sur (O, F) et telle que, pour tout $F \in F$:

$$J(A, F) = \int_F I(A, \xi) d\lambda(\xi)$$

si les conditions de la proposition sont vérifiées, le théorème de Radon-Nikodym [25] nous permet d'affirmer dans ce cas l'existence et l'unicité de la variable aléatoire réelle positive $I(A, \cdot)$ définie à une λ -équivalence près sur (O, F) , vérifiant la relation demandée. Inversement, si l'on se donne $I(A, \cdot)$ telle que pour tout $F \in F$

$$J(A, F) = \int_F I(A, \xi) d\lambda(\xi)$$

il est clair que les groupes disjoints d'observateurs sont indépendants pour l'observation de l'événement $A \in S$. Ce résultat est à rapprocher de celui de J. KAMPE DE FERIET dans [14], nous reviendrons sur cette comparaison au paragraphe V.1.

Soient A_1 et A_2 deux éléments de S tels que $A_1 \subseteq A_2$. Nous avons : $I(A_1, \cdot) \geq I(A_2, \cdot)$ λ -presque surement. En effet $A_1 \subseteq A_2$ implique par définition pour tout $F \in \mathcal{F}$: $J(A_1, F) \geq J(A_2, F)$ c'est-à-dire :

$$\int_F (I(A_1, \xi) - I(A_2, \xi)) d\lambda(\xi) \geq 0$$

ce qui entraîne :

$$I(A_1, \cdot) - I(A_2, \cdot) \geq 0 \quad \lambda\text{-presque surement.}$$

On peut donc associer à chaque couple d'événements $(A_1, A_2) \in S \times S$ un ensemble λ -négligeable de $\mathcal{F}$, noté $N(A_1, A_2)$, tel que pour tout $\xi \in \mathcal{O} - N(A_1, A_2)$, on ait :

$$\begin{aligned} \text{III.3.2.-} \quad & I(A_i, \xi) \geq 0 \quad i = 1, 2 \\ & A_1 \subseteq A_2 \Rightarrow I(A_1, \xi) \geq I(A_2, \xi). \end{aligned}$$

Comparons les propositions III.3.1. et III.1.1. On voit d'une part, que dans le cas fini, l'application $I_\xi(A)$ est parfaitement déterminée pour tout $\xi \in \mathcal{O}$ et pour tout $A \in S$ à partir de J , alors que dans le cas général on affirme simplement l'existence de l'application $I(A, \cdot)$ pour $A \in S$; d'autre part, pour tout $\xi \in \mathcal{O}$, $I_\xi(\cdot)$ est une information au sens de J. KAMPE DE FERIET sur (Ω, S) ce que nous ne pouvons pas affirmer pour $I(\cdot, \xi)$ dans le cas général. En effet, les relations III.3.2. ne sont vraies qu'à un

ensemble λ -négligeable près; or les ensembles $\bigcup_{A_1 \in S} N(A_1, A_2)$ pourraient ne pas

être λ -négligeables. Nous nous trouvons devant un problème semblable à celui qui se pose pour la définition des probabilités conditionnelles régulières [25] c'est-à-dire que, dans le cas général, nous ne pouvons pas affirmer qu'il existe pour chaque $A_1 \in S$ une version de la classe d'équivalence de $I(A_1, \cdot)$ qui soit telle que l'application I définisse une décomposition régulière de J au sens de la définition ci-après.

III.3.3.- DEFINITION.

Etant donné une information J sur $(\Omega \times 0, S \times F)$ et λ une mesure de probabilité sur $(0, F)$ telles que $N \subseteq F_0$; nous dirons qu'il existe une décomposition régulière de J s'il existe une application I_1 de $(S \times 0)$ dans $\bar{\mathbb{R}}^+$ telle que

1) Pour tout $\xi \in 0$, $I_1(\cdot, \xi)$ est une information au sens

I.1.1. sur (Ω, S) .

2) Pour tout $A \in S$, $I_1(A, \cdot)$ est mesurable sur $(0, F)$.

3) Pour tout $(A, F) \in S \times F$:

$$J(A, F) = \int_F I_1(A, \xi) d\lambda(\xi).$$

III.3.4.- COROLLAIRE.

S'il existe une décomposition régulière de J tous les groupes disjoints d'observateurs de $(0, F)$ sont σ -indépendants en information pour S et les applications I (de la proposition III.3.1.) et I_1 (de la définition III.3.3.) vérifient pour tout $A \in S$:

$$I(A, \cdot) = I_1(A, \cdot) \quad \lambda\text{-presque sûrement.}$$

Nous n'étudierons pas plus en détail ce problème de l'existence d'une décomposition régulière de J car il ne se pose pas dans les applications que nous envisageons par la suite.

Dans notre note [7] nous avons énoncé les axiomes de définition de J en supposant $F_0 = \{\emptyset\}$; il est clair que les raisonnements utilisés dans ce paragraphe nécessitent une définition plus générale de la famille F_0 des groupes d'observateurs qui n'apportent pas d'information ; c'est une des raisons pour lesquelles nous avons été amenés à modifier légèrement les axiomes 1 et 2 de [7] pour leur donner la forme qu'ils ont au paragraphe I.2.8.

CHAPITRE IV

CONSTRUCTION D'UNE INFORMATION APPORTEE PAR
UN COUPLE D'OBSERVATEURS.

Pour déterminer une information $J(A, F)$ apportée sur un événement $A \in S$ par un groupe $F \in F$ d'observateurs, suivant les méthodes proposées dans le chapitre III, il suffit de connaître les informations $J(A, \{\xi\})$ et $J(A, \{\xi_1, \xi_2\})$ apportées sur cet événement par les observateurs ξ et les couples d'observateurs (ξ_1, ξ_2) du groupe F . C'est pourquoi nous nous proposons de définir à partir d'une hypothèse particulière une information J sur $(\Omega \times O, S \times F)$ quand l'ensemble O est composé de deux observateurs ξ_a et ξ_b .

Nous nous restreignons, dans ce chapitre, au cas d'une sous-tribu T de S comportant un nombre fini d'éléments.

Les principaux résultats de cette partie ont été exposés dans notre note [23] : "Information moyenne dans une épreuve statistique (cas fini)".

IV.1. - NOTATIONS - HYPOTHESES.

Avant de définir les notations et les hypothèses spécifiques de cette partie, replaçons ce travail dans le cadre général des chapitres précédents. Rappelons d'abord les grandes lignes de la construction proposée dans le chapitre II d'une information J sur un espace produit, en l'appliquant au cas :

$$O = \{\xi_a, \xi_b\}$$

$$F = \{\emptyset, \{\xi_a\}, \{\xi_b\}, O\}.$$

Soient $(\Omega_a, S_a), (\Omega_b, S_b)$ les espaces mesurables associés respectivement aux observateurs ξ_a, ξ_b et $(\hat{\Omega}, \hat{S})$ l'espace mesurable produit défini par :

$$\hat{\Omega} = \Omega_a \times \Omega_b$$

$$\hat{S} = S_a \otimes S_b.$$

Conformément aux notations précédemment définies (I.1.3) pour tout $\hat{C} \in \hat{S}$ nous notons C_a et C_b les projections de $\hat{C}$ respectivement sur Ω_a et Ω_b , $\hat{C}_a$ et $\hat{C}_b$ les cylindres de $\hat{S}_a$ et $\hat{S}_b$ de base C_a et C_b . Soit I une information sur $(\hat{\Omega}, \hat{S})$, définissons l'application J sur $(\hat{\Omega} \times O, \hat{S} \times F)$ par : $\forall \hat{C} \in \hat{S}$

$$J(\hat{C}, \emptyset) = 0$$

$$J(\hat{C}, \{\xi_a\}) = I(\hat{C}_a)$$

$$J(\hat{C}, \{\xi_b\}) = I(\hat{C}_b)$$

$$J(\hat{C}, 0) = I(\hat{C}) .$$

Comme nous l'avons montré au chapitre II, il est facile de vérifier que l'application J , ainsi définie à partir de I , est une information sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ au sens I.2.2 - I.2.6.

Afin de simplifier les notations et de nous placer dans un cadre un peu plus général nous noterons (Ω, S) l'espace des événements que nous employons. Cet espace peut-être identifié à l'espace produit $(\hat{\Omega}, \hat{S})$ ou être différent de celui-ci.

Il est essentiel de remarquer pour cette identification

que les sous- σ -algèbres de Boole A et B de S que nous employons, tiennent le rôle de $\hat{S}_a$ et $\hat{S}_b$ et que, comme ces dernières, elles sont supposées M -indépendantes et spécifiquement liées aux observateurs ξ_a et ξ_b .

IV.1.1. - NOTATIONS.

Soient Ω un ensemble quelconque, S une σ -algèbre de parties de Ω et P une pondération sur (Ω, S) , c'est-à-dire une mesure positive, finie, de masse totale égale à 1, qui pourra être, le cas échéant, une mesure de fréquence ou de probabilité.

Soient A et B deux sous-algèbres de S comportant un nombre fini d'éléments, supposées M -indépendantes (I.1.1.) ; nous avons donc :

$$(A, B) \in A \times B, A \neq \emptyset, B \neq \emptyset \implies A \cap B \neq \emptyset.$$

Nous écrirons indifféremment $A.B$ ou $A \cap B$ l'intersection de deux événements.

Notons T la sous-tribu de S engendrée par A et B (T comporte un nombre fini d'éléments). Quand nous utilisons les espaces mesurés (Ω, A, P) , (Ω, B, P) , (Ω, T, P) nous nous référons aux restrictions aux tribus A, B, T de la mesure P définie sur (Ω, S) . Nous supposons l'existence de tous les ensembles et de toutes les partitions de ces ensembles dont nous avons besoin. En particulier, la mesure P peut prendre sur A, B, T toutes les valeurs possibles compatibles avec les axiomes de définition d'une mesure et les restrictions que nous imposons.

Pour simplifier les énoncés et les démonstrations, nous supposons, de plus, que :

$$(A, B) \in A \times B, A \neq \emptyset, B \neq \emptyset \implies P(A) > 0 \text{ et } P(B) > 0$$

Pour $A \in A$, $A \neq \emptyset$ et $A \neq \Omega$ nous avons donc $0 < P(A) < 1$, de même, pour $B \in B$; notons, par contre, qu'il est possible d'obtenir $C \in T$, $C \neq \emptyset$ et $P(C) = 0$.

Nous notons $\Pi_A(A)$ (respectivement $\Pi_A = \Pi_A(\Omega)$) une partition finie A -mesurable de $A \in A$ (resp. de Ω) :

$$\begin{aligned} \cdot (A_1, A_2) \in \Pi_A(A) \times \Pi_A(A) &\implies \begin{cases} (A_1, A_2) \in A \times A \\ A_1 \neq \emptyset \text{ et } A_2 \neq \emptyset \\ A_1 \cap A_2 = \emptyset \text{ ou } A_1 = A_2 \end{cases} \\ \cdot \cup \{A' \mid A' \in \Pi_A(A)\} &= A \end{aligned}$$

Les partitions $\Pi_A(A)$ sont quelconques, en particulier Π_A peut être égale à la partition qui engendre la tribu finie A ou différente de cette partition.

Nous notons de même $\Pi_B(B)$ et $\Pi_B = \Pi_B(\Omega)$ des partitions finies B -mesurables de $B \in B$ et de Ω .

REMARQUES. -

Nous imposons donc essentiellement trois conditions aux espaces mesurés que nous allons employer :

- . les tribus A, B, T comportent un nombre fini d'éléments
- . sauf pour l'ensemble vide, tous les éléments de A et B sont de mesure strictement positive
- . les tribus A et B sont M -indépendantes :

$$(A, B) \in A \times B, \quad A \neq \emptyset, \quad B \neq \emptyset \implies A \cap B \neq \emptyset.$$

Ces restrictions, qui facilitent les démonstrations ultérieures, s'introduisent naturellement dans les études pratiques de statistique : les variables statistiques qualitatives ou quantitatives discrètes ne prennent qu'un nombre fini de valeurs et l'on regroupe en un nombre fini de classes les valeurs numériques prises par les variables statistiques continues ; si P_{ij} est la probabilité (ou la fréquence) correspondant à l'événement $A_i \cap B_j$, on suppose le plus souvent en statistique :

$$P_{i.} = P(A_i) = \sum_j P_{ij} > 0$$

$$P_{.j} = P(B_j) = \sum_i P_{ij} > 0$$

et seulement $P_{ij} \geq 0.$

IV.1.2.- HYPOTHESES.

L'information généralisée I au sens de J. KAMPE DE FERIET que nous cherchons à déterminer, est définie sur l'espace mesurable (Ω, T) et doit vérifier les axiomes 1, 2, 3 et les définitions que nous avons rappelés en I.1.1.

Dans une épreuve statistique, quand un événement C de la forme $C = A \cap B$, $A \in \mathcal{A}$ et $B \in \mathcal{B}$, se réalise, nous supposons que le premier observateur ξ_a perçoit l'événement A et le second observateur ξ_b l'événement B . Les différentes quantités $I(A/B_1)$, $I(A/B_2)$, ..., $I(A/B_n)$ représentent l'information apportée par l'observation de A par ξ_a quand on suppose connue l'information apportée par l'observateur ξ_b dans les différentes hypothèses $B_i \in \Pi_B$ possibles. La quantité

$$\sum \{P(B) I(A/B) \mid B \in \Pi_B\}$$

représente l'information moyenne apportée par A quand on réalise un grand nombre de fois cette épreuve. A notre sens, cette information moyenne doit être égale à $I(A)$.

Par exemple, si Ω est un ensemble de 100 étudiants répartis en deux groupes : B_1 les 80 étudiants à plein temps, B_2 les 20 auditeurs salariés étudiant à temps partiel, l'événement A étant la réussite à l'examen, nous pensons que l'information apportée sur un individu quelconque par le succès à l'examen est la moyenne pondérée des informations apportées par la réussite à l'examen dans chacun des groupes :

$$I(A) = \frac{80}{100} I(A/B_1) + \frac{20}{100} I(A/B_2) .$$

Dans cet exemple, ξ_a est l'observateur lié au caractère : réussite ou non réussite à l'examen; ξ_b est l'observateur qui perçoit le caractère : statut de l'étudiant (plein temps ou temps partiel).

H.1.- HYPOTHESE 1.

Pour tout $(A,B) \in \mathcal{A} \times \mathcal{B}$ et quelles que soient les partitions finies Π_A et Π_B respectivement $\mathcal{A}$ -mesurables et $\mathcal{B}$ -mesurables de Ω , nous avons

$$I(A) = \sum \{P(B') I(A/B') | B' \in \Pi_B\}$$

$$I(B) = \sum \{P(A') I(B/A') | A' \in \Pi_A\}$$

Ces sommes ont toujours un sens car les hypothèses que nous avons imposées à $(\Omega, \mathcal{A}, P)$ et $(\Omega, \mathcal{B}, P)$ évitent l'introduction de termes indéfinis de la forme

$$P(B') I(A/B') = 0 . + \infty$$

ou

$$P(A') I(B/A') = 0 . + \infty$$

En effet, si les quantités $I(A/B')$ et $I(B/A')$ appartiennent à $\bar{\mathbb{R}}^+$; $A' \in \Pi_A$ entraîne $A' \in \mathcal{A}$ et $A' \neq \emptyset$ c'est-à-dire $P(A') > 0$ et de même $B \in \Pi_B$ implique $P(B') > 0$.

Si les tribus $\mathcal{A}$ et $\mathcal{B}$ ne sont pas supposées M -indépendantes et, en particulier, si $\mathcal{A} = \mathcal{B}$, l'hypothèse 1 et les axiomes de l'information généralisée impliquent alors $I(A) = +0$ ou $+\infty$, pour tout $A \in \mathcal{A}$. Dans ce cadre l'hypothèse 1 n'aurait évidemment aucun intérêt.

Dans l'hypothèse 1, nous nous intéressons seulement aux ensembles $A \in \mathcal{A}$, $B \in \mathcal{B}$ et aux éléments $C \in \mathcal{T}$ de la forme $C = A \cap B$ pour cette raison, il est commode de déterminer d'abord l'information I sur ces ensembles de $\mathcal{T}$ telle manière que I vérifie l'hypothèse 1, et de prolonger ensuite I

à l'ensemble T tout entier. Nous allons donc, dans l'hypothèse suivante, nous restreindre à l'étude de $I(A,B)$ et supposer que $I(A,B)$ est une fonction continue de $P(A)$, $P(B)$, $P(A.B)$.

$$(A,B) \in A \times B : I(A,B) = F(P(A), P(B), P(A.B))$$

Notons que les informations du type WIENER SHANNON sont bien de cette forme :

$$I(A,B) = C \cdot \text{Log} \frac{1}{P(A.B)} \quad (C > 0)$$

mais qu'elles ne vérifient pas l'hypothèse 1 en règle générale :

$$I(A) = C \cdot \text{Log} \frac{1}{P(A)} \neq \sum \{ P(B) \cdot C \cdot \text{Log} \frac{P(B)}{P(A.B)} \mid B \in \pi_B \}$$

Définissons l'ensemble de définition de F en explicitant les contraintes suivantes (avec les notations de l'axiome 3 de I.1.1) :

$$\begin{aligned} 0 \leq P(A^{(i)}) &\leq 1 & i = 0,1 \\ 0 \leq P(B^{(j)}) &\leq 1 & j = 0,1 \\ 0 \leq P(A^{(i)} \cdot B^{(j)}) &\leq \text{Inf}[P(A^{(i)}), P(B^{(j)})] & i,j = 0,1 \end{aligned}$$

Posons $x = P(A)$, $y = P(B)$, $z = P(A.B)$, il vient :

$$\begin{aligned} 0 \leq x = P(A) &\leq 1 \\ 0 \leq y = P(B) &\leq 1 \\ \text{Sup}[0, x+y-1] \leq z = P(A.B) &\leq \text{Inf}[x, y] \end{aligned}$$

Donc :

$$E = \{(x,y,z) \in [0,1]^3 \mid \text{Sup}[0, x+y-1] \leq z \leq \text{Inf}[x, y]\}$$

est l'ensemble de définition de F . Pour tout triplet $(x,y,z) \in E$, il existe $A \in \mathcal{A}$, $B \in \mathcal{B}$ et une pondération P sur (Ω, T) qui vérifient $P(A) = x$, $P(B) = y$, $P(A.B) = z$ et inversement.

Nous utiliserons aussi l'ensemble :

$$E^* = E \cap \{(x,y,z) \mid (x,y) \in]0,1]^2\}$$

Tous les points de $E - E^*$ sont de la forme $(0,y,0)$ avec $y \in [0,1]$ ou $(x,0,0)$ avec $x \in [0,1]$. En particulier, si $A.B = \emptyset$ on sait que $A = \emptyset$ et/ou $B = \emptyset$, c'est-à-dire que le triplet $(P(A), P(B), P(A.B)) \in E - E^*$; dans ce cas, nous savons que $I(A.B) = +\infty$. Il est donc nécessaire que $F(x,y,z)$ soit égal à $+\infty$ sur $E - E^*$, en règle générale, il se pourrait que $F(x,y,z) = +\infty$ sur certains points de E (c'est-à-dire : $I(A.B) = +\infty$ avec $A.B \neq \emptyset$ et même $P(A.B) > 0$). Nous allons exclure cette possibilité dans l'hypothèse ci-dessous :

H.2.- HYPOTHESE 2.

Pour $(A,B) \in A \times B$ l'information apportée par l'événement $A.B$ est une fonction de $P(A), P(B), P(A.B)$.

Précisément :

$$I(A.B) = F(P(A), P(B), P(A.B))$$

où F est une application de E dans $\bar{\mathbb{R}}^+$, finie et continue sur E^ par rapport à chaque variable, et égale à $+\infty$ sur $E - E^*$*

$$E = \{(x,y,z) \in [0,1]^3 \mid \sup(0, x+y-1) \leq z \leq \inf(x,y)\}$$

$$E^* = E \cap \{(x,y,z) \in]0,1]^2 \times [0,1]\}.$$

L'application F , définie dans cette hypothèse et utilisée dans ce chapitre, n'a évidemment aucun rapport avec celle de la loi de composition (I.1.1 - Axiome 4). Rappelons que les informations I que nous utilisons ne sont pas supposées composables.

Il est clair que, si les sous-algèbres A et B de T ne sont pas supposées M -indépendantes, cette hypothèse n'a aucun sens.

IV.2.- RESULTATS PRELIMINAIRES.CONSEQUENCES DE L'HYPOTHESE 1.IV.2.1.- PROPOSITION.

Si l'information I définie sur l'espace mesuré $(\Omega, \mathcal{T}, P)$ vérifie l'hypothèse 1, alors pour $A \in \mathcal{A}$, nous avons :

$$I(A) = +\infty \iff \forall B \in \mathcal{B} : I(A/B) = +\infty \iff \exists B \in \mathcal{B} : I(A/B) = +\infty$$

$$I(A) < +\infty \iff \forall B \in \mathcal{B} : I(A/B) < +\infty \iff \exists B \in \mathcal{B} : I(A/B) < +\infty$$

et des relations identiques pour $B \in \mathcal{B}$.

Une des conséquences de la définition de l'information conditionnelle est que :

$$I(A) = +\infty \iff \forall B \in \mathcal{B} : I(A/B) = +\infty \implies \exists B \in \mathcal{B} : I(A/B) = +\infty$$

Supposons maintenant qu'il existe $B \in \mathcal{B}$ tel que $I(A/B) = +\infty$

- . si $I(B) = +\infty$ alors $I(A/B) = I(A) = +\infty$
- . si $I(B) < +\infty$ nous avons donc $B \neq \emptyset$, $P(B) > 0$ et l'on déduit $I(A) = +\infty$ de l'hypothèse 1 :

$$I(A) = P(B) I(A/B) + P(B^c) I(A/B^c)$$

Nous avons ainsi montré la première série d'équivalences; la seconde série s'en déduit immédiatement.

IV.2.2.- PROPOSITION.

Une condition nécessaire et suffisante pour qu'une information I définie sur l'espace mesuré $(\cdot, \cdot, P)$ vérifie l'hypothèse 1 est que :

- . pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$
- . pour tout $(A_1, B_1) \in \mathcal{A} \times \mathcal{B}$, $A_1 \neq \emptyset$ et $B_1 \neq \emptyset$
- . quelles que soient les partitions finies $\Pi_A(A_1)$ et $\Pi_B(B_1)$

respectivement A -mesurable de A_1 et B -mesurable de B_1
on ait les relations :

$$P(B_1) I(A/B_1) = \sum \{P(B') I(A/B') | B' \in \pi_B(B_1)\}$$

$$P(A_1) I(B/A_1) = \sum \{P(A') I(B/A') | A' \in \pi_A(A_1)\}$$

On montre que la condition est suffisante en prenant $(A_1, B_1) = (\Omega, \Omega)$. Il est facile de voir que cette condition est nécessaire; par exemple, l'hypothèse 1 implique les deux décompositions suivantes de $I(A)$:

$$I(A) = P(B_1^C) I(A/B_1^C) + P(B_1) I(A/B_1)$$

$$I(A) = P(B_1^C) I(A/B_1^C) + \sum \{P(B') I(A/B') | B' \in \pi_B(B_1)\}$$

Si $I(A)$ est finie, tous les termes $I(A/B')$ le sont aussi et la relation est vérifiée si $I(A) = +\infty$ tous les termes $I(A/B')$ sont égaux à $+\infty$ et la relation est vérifiée.

IV.2.3.- PROPOSITION.

Une condition nécessaire et suffisante pour qu'une information I définie sur l'espace mesuré (Ω, T, P) vérifie l'hypothèse 1 est que

- . pour tout $(A, B) \in A \times B$
- . pour tout couple d'ensembles disjoints A_1, A_2 de A et B_1, B_2 de B

$$A_1 \neq \emptyset, A_2 \neq \emptyset, B_1 \neq \emptyset, B_2 \neq \emptyset$$

on ait :

$$P(B_1 \cup B_2) I(A/B_1 \cup B_2) = P(B_1) I(A/B_1) + P(B_2) I(A/B_2)$$

$$P(A_1 \cup A_2) I(B/A_1 \cup A_2) = P(A_1) I(B/A_1) + P(A_2) I(B/A_2)$$

La condition est évidemment nécessaire, elle est aussi suffisante car il suffit de l'appliquer de proche en proche un nombre fini de fois pour réobtenir la proposition IV.2.2.

IV.2.4.- PROPOSITION.

Soient A_1 et A_2 deux ensembles disjoints de A ; s'il existe $B \in \mathcal{B}$ tel que : $I(B/A_1) = I(B/A_2)$ alors on a aussi $I(B/A_1) = I(B/A_2) = I(B/A_1 \cup A_2)$. De même, s'il existe $(A, B_1, B_2) \in A \times \mathcal{B} \times \mathcal{B}$ tel que $I(A/B_1) = I(A/B_2)$ et $B_1 \cdot B_2 = \emptyset$ alors $I(A/B_2) = I(A/B_1 \cup B_2)$.

C'est une conséquence immédiate de la proposition précédente que l'on pourrait d'ailleurs généraliser au cas d'un nombre fini quelconque d'ensembles disjoints.

CONSEQUENCES DE L'HYPOTHESE 2.

Soit l'espace mesuré (Ω, T, P) , nous avons défini l'ensemble E des valeurs que peut prendre le triplet $(P(A), P(B), P(A.B))$ pour $(A, B) \in A \times \mathcal{B}$; nous allons maintenant caractériser les fonctions F qui peuvent définir une information par la formule :

$$I(A.B) = F(P(A), P(B), P(A.B)) .$$

IV.2.5.- PROPOSITION.

Soit une application F définie sur E à valeurs dans $\bar{\mathbb{R}}^+$.

$$E = \{(x, y, z) \in [0, 1]^3 \mid \sup(0, x+y-1) \leq z \leq \inf(x, y)\}$$

Posons, pour $(A, B) \in A \times \mathcal{B}$:

$$I(A.B) = F(P(A), P(B), P(A.B)) .$$

Une condition nécessaire et suffisante pour que I définisse sur les éléments de T de la forme $A.B$, $(A, B) \in A \times \mathcal{B}$, une information qui vérifie les axiomes A.1 et A.2 de I.1.1. est que :

$$* F(1,1,1) = 0$$

$$* F(0,0,0) = F(x,0,0) = F(0,y,0) = +\infty \quad \forall (x,y) \in [0,1]^2$$

$$* \forall (x,y,z) \in E, \forall (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in [0,1]^3 \text{ tels que :}$$

$$0 \leq \varepsilon_3 \leq \varepsilon_1 + \varepsilon_2$$

$$\text{et } (x+\varepsilon_1, y+\varepsilon_2, z+\varepsilon_3) \in E$$

on ait :

$$F(x+\varepsilon_1, y+\varepsilon_2, z+\varepsilon_3) \leq F(x,y,z)$$

Ces conditions sont nécessaires :

$$* I(\Omega) = F(1,1,1) = 0$$

$$* I(\emptyset) = I(A, \emptyset) = I(\emptyset, B) = +\infty \quad \text{pour tout } (A,B) \in A \times B$$

$$\text{donc } F(0,0,0) = F(x,0,0) = F(0,y,0) = +\infty$$

$$\text{pour tout } (x,y) \in [0,1]^2$$

Soient $(A,B) \in A \times B$ et $(A_1, B_1) \in A \times B$ tels que $A_1 \supset A$ et $B_1 \supset B$; nous avons

$$A_1 \cdot B_1 \supset A \cdot B \quad \text{et} \quad I(A_1 \cdot B_1) \leq I(A \cdot B)$$

ce qui s'écrit : $F(x+\varepsilon_1, y+\varepsilon_2, z+\varepsilon_3) \leq F(x,y,z)$ avec les notations $x = P(A)$, $y = P(B)$, $z = P(A \cdot B)$

$$x+\varepsilon_1 = P(A_1), \quad y+\varepsilon_2 = P(B_1), \quad z+\varepsilon_3 = P(A_1 \cdot B_1)$$

Notons que : $\varepsilon_3 \leq \varepsilon_1 + \varepsilon_2$ car :

$$\varepsilon_3 = P(A_1 \cdot B_1) - P(A \cdot B) = P(A_1 \cdot B_1) - P(A_1 \cdot B) + P(A_1 \cdot B) - P(A \cdot B)$$

$$\text{et} \quad P(A_1.B_1) - P(A_1.B) = P(A_1.(B_1-B)) \leq P(B_1-B) = \epsilon_2$$

$$P(A_1.B) - P(A.B) = P((A_1-A).B) \leq P(A_1-A) = \epsilon_1$$

Il est facile de voir en faisant varier (A,B) , (A_1,B_1) que ces relations doivent être vérifiées $\forall (x,y,z) \in E$, $\forall (\epsilon_1, \epsilon_2, \epsilon_3) \in [0,1]^3$ tels que $(x+\epsilon_1, y+\epsilon_2, z+\epsilon_3) \in E$ et $\epsilon_3 \leq \epsilon_1 + \epsilon_2$.

Ces conditions sont aussi suffisantes, l'axiome A.1 est vérifié car, pour tout $(x,y,z) \in E$ en posant $\epsilon_1 = 1-x$, $\epsilon_2 = 1-y$, $\epsilon_3 = 1-z$ on déduit $\epsilon_3 \leq \epsilon_1 + \epsilon_2$ de $z \geq x+y-1$ et $0 = F(1,1,1) \leq F(x,y,z)$ c'est-à-dire que $I(A.B) \in \mathbb{R}^+$ pour tout $(A,B) \in A \times B$. Les valeurs universelles sont définies; il reste à montrer que, pour $(A,B) \in A \times B$, $(A_1,B_1) \in A \times B$ tels que $A.B \subset A_1.B_1$ on a $I(A.B) \geq I(A_1.B_1)$. Pour cela, montrons d'abord que $\emptyset \neq A.B \subset A_1.B_1$ entraîne obligatoirement $A \subset A_1$ et $B \subset B_1$. En effet, définissons, par exemple, $A' = A - A_1$, nous avons :

$$* A' \subset A \implies A'.B \subset A.B \subset A_1.B_1 \subset A_1 \quad \text{d'une part,}$$

$$* A' \subset A_1^c \implies A'.B \subset A_1^c \quad \text{d'autre part}$$

c'est-à-dire que $A'.B = \emptyset$ et, comme $B \neq \emptyset$, nous devons conclure que $A' = \emptyset$ par suite des hypothèses du paragraphe I.1. (M-indépendance de A et B) $A \subset A_1$ et de même $B \subset B_1$. Par le raisonnement inverse de celui que nous avons employé ci-dessus pour montrer la condition nécessaire, on trouve que :

$$F(P(A_1), P(B_1), P(A_1.B_1)) \leq F(P(A), P(B), P(A.B))$$

$$\text{c'est-à-dire :} \quad I(A_1.B_1) \leq I(A.B)$$

La proposition est ainsi démontrée et l'on en déduit immédiatement :

IV.2.6.- PROPOSITION.

Une condition nécessaire et suffisante pour que l'application F de la proposition IV.2.5. définisse une information qui vérifie l'hypothèse 2 est : les conditions de la proposition IV.2.5. sont vérifiées; de plus F est continue par rapport à chaque variable et finie sur E^* .

$$E^* = E \cap \{(x,y,z) \in]0,1]^2 \times [0,1]\}$$

L'information est ainsi définie pour tous les éléments des tribus A et B : $(A,B) \in (A \times B)$

$$I(A) = I(A,\Omega) = F(P(A),1,P(A)) = f_1(P(A))$$

$$I(B) = I(\Omega,B) = F(1,P(B),P(B)) = f_2(P(B))$$

IV.2.7.- PROPOSITION.

Une condition nécessaire et suffisante pour que la relation $I(A) = f_1(P(A))$ (respectivement : $I(B) = f_2(P(B))$) définisse une information sur la tribu A (resp. B) est que f_1 (resp. f_2) soit une application non croissante de $[0,1]$ dans $\bar{\mathbb{R}}^+$ telle que $f_1(0) = +\infty$ et $f_1(1) = 0$ (resp. $f_2(0) = +\infty$ et $f_2(1) = 0$). Si l'on veut que l'hypothèse 2 soit vérifiée, il faut, en plus que f_1 (resp. f_2) soit une application finie et continue sur $]0,1[$.

Ce résultat est évident, il conduit naturellement à introduire la notion d'information mutuelle. Soit $(A,B) \in A \times B$ on appelle information mutuelle (I.1.1 - c) entre A et B la quantité :

$$I(A;B) = I(A) + I(B) - I(A.B) \quad \text{si } I(A).I(B) < +\infty$$

$$I(A;B) = 0 \quad \text{sinon.}$$

L'information sur les éléments de T de la forme $A.B$, $A \in A$ et $B \in B$ s'écrit alors comme la somme de trois termes; une information sur A , une information sur B et l'information mutuelle entre A et B .

$$I(A.B) = I(A) + I(B) - I(A;B)$$

Comme nous le verrons dans la suite, il est assez naturel d'envisager le cas où $I(A)$ est une information arbitraire sur A , $I(B)$ une information arbitraire sur B ; dans ce cas $I(A.B)$ est défini à partir de $I(A;B)$.

Nous avons :

$$I(A.B) = F(P(A), P(B), P(A.B))$$

$$I(A) = F(P(A), 1, P(A)) = f_1(P(A))$$

$$I(B) = F(1, P(B), P(B)) = f_2(P(B))$$

$$I(A;B) = f_1(P(A)) + f_2(P(B)) - F(P(A), P(B), P(A.B))$$

définissons l'application θ de E à valeurs dans $\bar{\mathbb{R}}$ par :

$$\theta(x,y,z) = f_1(x) + f_2(y) - F(x,y,z) \quad \text{si } f_1(x) \cdot f_2(y) < +\infty$$

et

$$\theta(x,y,z) = 0 \quad \text{si } f_1(x) \cdot f_2(y) = +\infty$$

il vient :

$$I(A;B) = \theta(P(A), P(B), P(A.B))$$

Il est clair que la donnée de F entraîne celles de f_1 , f_2 et θ ; inversement, on peut définir F à partir de f_1 , f_2 et θ .

IV.2.8.- PROPOSITION.

Soit I une application définie pour $(A,B) \in A \times B$ par :

$$I(A) = f_1(P(A))$$

$$I(B) = f_2(P(B))$$

$$I(A,B) = f_1(P(A)) + f_2(P(B)) - \theta(P(A), P(B), P(A.B))$$

$$\text{si } f_1(P(A)) + f_2(P(B)) < +\infty$$

$$I(A,B) = +\infty \text{ sinon.}$$

Une condition nécessaire et suffisante pour que I définisse sur les éléments de la forme $A \cap B$, $(A,B) \in A \times B$ une information qui vérifie l'hypothèse 2 et les axiomes A.1, A.2 de I.1.1. est :

- a) f_1 et f_2 sont des applications non croissantes de $[0,1]$ dans $\mathbb{R}^+$, finies et continues sur $]0,1]$ et telles que :

$$f_1(0) = f_2(0) = +\infty$$

$$f_1(1) = f_2(1) = 0$$

- b) θ est une application finie et continue par rapport à chaque variable, de E^* dans $\mathbb{R}$ telle que :

$$\theta(1,1,1) = \theta(x,1,x) = \theta(1,y,y) = 0 \quad \forall (x,y) \in [0,1]^2$$

Sur $E-E^*$, $\theta(x,y,z)$ peut être prolongé arbitrairement dans $\mathbb{R}$ par exemple en posant :

$$\theta(x,0,0) = \theta(0,y,0) = 0 \quad \forall (x,y) \in [0,1]^2$$

ce qui nous permet d'écrire

$$F(x,y,z) = f_1(x) + f_2(y) - \theta(x,y,z)$$

quelque soit $(x,y,z) \in E$ (et d'énoncer la condition c sur E et non seulement sur E^*).

c) pour $(x,y,z) \in E$, $(x',y,z') \in E$ tels que
 $x' - x \geq z' - z \geq 0$

$$f_1(x) - f_1(x') \geq \theta(x,y,z) - \theta(x',y,z')$$

de même, pour $(x,y,z) \in E$, $(x,y',z') \in E$ tels que
 $y' - y \geq z' - z \geq 0$

$$f_2(y) - f_2(y') \geq \theta(x,y,z) - \theta(x,y',z')$$

L'hypothèse 2 implique que $F(x,y,z)$, donc $f_1(x)$ et $f_2(y)$, sont finis et continus par rapport à chaque variable sur E^* et que, sur $E-E^*$:

$$F(x,y,z) = +\infty, \quad f_1(x) = +\infty \quad \text{et/ou}$$

$$F(x,y,z) = +\infty, \quad f_2(y) = +\infty$$

Cela entraîne que θ est une application finie et continue par rapport à chaque variable de E^* dans $\mathbb{R}$ et que les conditions a et b sont nécessaires. Par définition, nous avons $F(x,1,x) = f_1(x)$; on déduit la condition a de la proposition IV.2.7. ; on en déduit de plus que $f_2(1) - \theta(x,1,x) = 0$ pour tout $x \in [0,1]$, c'est-à-dire $\theta(x,1,x) = 0$ pour tout $x \in [0,1]$; de même $\theta(1,y,y) = \theta(1,1,1) = 0$ pour tout $y \in [0,1]$.

Prolongeons $\theta(x,y,z)$ par zéro sur $E-E^*$ comme cela est indiqué dans l'énoncé.

La condition c est nécessaire ; en effet, pour tout $(x,y,z) \in E$ et $(x',y,z') \in E$ tels que $x' - x \geq z' - z \geq 0$, on peut

définir $\epsilon_1 = x' - x$, $\epsilon_2 = 0$, $\epsilon_3 = z' - z$ et $(\epsilon_1, \epsilon_2, \epsilon_3) \leq E$,
 $0 \leq \epsilon_3 \leq \epsilon_1$. En appliquant la proposition IV.2.5., il vient :

$$F(x+\epsilon_1, y+\epsilon_2, z+\epsilon_3) \leq F(x, y, z)$$

c'est-à-dire

$$F(x', y, z') \leq F(x, y, z).$$

On en déduit :

$$f_1(x) - f_1(x') \geq \theta(x, y, z) - \theta(x', y, z')$$

On démontrerait de même la seconde partie de la condition c.

Les conditions a, b, c sont aussi suffisantes : il est clair que $F(1, 1, 1) = 0$ et que $F(0, 0, 0) = F(x, 0, 0) = F(0, y, 0) = +\infty$ pour tout $(x, y) \in [0, 1]^2$. En reprenant le raisonnement du paragraphe précédent, on obtient très facilement la troisième condition de la proposition IV.2.5. . En effet, pour tout $(x, y, z) \in E$, $(\epsilon_1, \epsilon_2, \epsilon_3) \in [0, 1]^3$ tels que :

$$0 \leq \epsilon_3 \leq \epsilon_1 + \epsilon_2$$

et

$$(x+\epsilon_1, y+\epsilon_2, z+\epsilon_3) \in E$$

on peut toujours définir ϵ'_1 et ϵ'_2 vérifiant $\epsilon_3 = \epsilon'_1 + \epsilon'_2$ et $0 \leq \epsilon'_1 \leq \epsilon_1$, $0 \leq \epsilon'_2 \leq \epsilon_2$; en appliquant c à

$$(x, y, z) \in E \text{ et } (x+\epsilon_1, y, z+\epsilon'_1)$$

d'une part, et à

$$(x+\epsilon_1, y, z+\epsilon'_1) \text{ et } (x+\epsilon_1, y+\epsilon_2, z+\epsilon'_1+\epsilon'_2)$$

d'autre part. Il vient :

$$F(x, y, z) \geq F(x+\epsilon_1, y, z+\epsilon'_1) \geq F(x+\epsilon_1, y+\epsilon_2, z+\epsilon_3)$$

C'est le résultat demandé. Nous sommes dans les conditions prévues dans

les propositions IV.2.5. et IV.2.6., c'est-à-dire que I ainsi définie vérifie les axiomes A.1 et A.2 ainsi que l'hypothèse 2.

Notons, de plus, que la décomposition ainsi obtenue est évidemment unique.

IV.3.- THEOREME

POSITION DU PROBLEME.

En rapprochant les conséquences des hypothèses 1 et 2, on définit un couple d'équations fonctionnelles qu'il est possible de résoudre sur E et dont la solution nous donne la forme générale de F . Les conséquences de l'hypothèse 1 montrent que, pour tout $A \in \mathcal{A}$, l'application $\mu_A : \mathcal{B} \rightarrow \bar{\mathbb{R}}^+$ définie par :

$$\mu_A(B) = P(B) I(A/B) \quad \text{si } P(B) > 0$$

$$\mu_A(B) = 0 \quad \text{sinon}$$

est une fonction d'ensemble simplement additive sur $\mathcal{B}$. De même, pour $B \in \mathcal{B}$ fixée, l'application $\mu_B : \mathcal{A} \rightarrow \bar{\mathbb{R}}^+$ définie par

$$\mu_B(A) = P(A) I(B/A) \quad \text{si } P(A) > 0$$

$$\mu_B(A) = 0 \quad \text{si } P(A) = 0$$

est une fonction simplement additive sur $\mathcal{A}$.

Les tribus $\mathcal{A}$ et $\mathcal{B}$ étant finies, μ_B et μ_A sont des mesures positives respectivement sur $(\Omega, \mathcal{A})$ et $(\Omega, \mathcal{B})$. Soient A_1 et A_2 des ensembles disjoints de $\mathcal{A}$; pour $B \in \mathcal{B}$ fixé, nous avons :

$$\mu_B(A_1 \cup A_2) = \mu_B(A_1) + \mu_B(A_2)$$

c'est-à-dire :

$$\text{IV.3.1.- } P(A_1 \cup A_2) I(B/A_1 \cup A_2) = P(A_1) I(B/A_1) + P(A_2) I(B/A_2)$$

$$\text{si } A_1 \neq \emptyset \text{ et } A_2 \neq \emptyset$$

Introduisons les notations de l'hypothèse 2 :

$$I(A.B) = F(P(A), P(B), P(A.B))$$

$$x_1 = P(A_1), x_2 = P(A_2), x_1 + x_2 = P(A_1 \cup A_2)$$

$$y = P(B)$$

$$z_1 = P(A_1.B), z_2 = P(A_2.B), z_1 + z_2 = P((A_1 \cup A_2).B)$$

. Si $I(A) < +\infty$ c'est-à-dire si $P(A) > 0$, on a :

$$I(B/A) = I(A.B) - I(A)$$

$$P(A) I(B/A) = P(A) [F(P(A), P(B), P(A.B)) - F(P(A), 1, P(A))]$$

Posons, dans ce cas :

$$\phi(P(A), P(B), P(A.B)) = P(A) I(B/A)$$

$$\text{IV.3.2.- } \phi(x, y, z) = x[F(x, y, z) - F(x, 1, x)]$$

La relation IV.3.1. s'écrit alors :

$$\text{IV.3.3.- } \phi(x_1 + x_2, y, z_1 + z_2) = \phi(x_1, y, z_1) + \phi(x_2, y, z_2)$$

. Si $I(A) = +\infty$ nous savons que l'hypothèse 2 implique $A = \emptyset$, c'est-à-dire $\mu_B(\emptyset) = 0$ par hypothèse.

$$\text{Posons donc } \phi(0, y, 0) = 0 \text{ pour } y \in [0, 1]$$

Rappelons que :

$$E = \{(x,y,z) \in [0,1]^3 \mid \sup(0,x+y-1) \leq z \leq \inf(x,y)\}$$

$$E^* = E \cap \{(x,y,z) \in]0,1]^2 \times [0,1]\}$$

et que l'ensemble $E - E^*$ est formé uniquement de points de la forme $(x,0,0)$ et $(0,y,0)$ pour $(x,y) \in [0,1]^2$.

Nous pouvons donc définir ϕ de la manière suivante :

. sur E^* l'application ϕ , définie par IV.3.2., est finie et continue

$$\begin{aligned} \text{. sur } E - E^* \quad & \left[\begin{array}{ll} - \text{ pour } y \in [0,1] & \phi(0,y,0) = 0 \\ - \text{ pour } x \in]0,1] & \phi(x,0,0) = +\infty \end{array} \right. \end{aligned}$$

Ainsi, ϕ vérifie la propriété IV.3.3. sur tous les points de E .

Notons, en particulier, que pour tout $(x,y,z) \in E^*$, nous avons $(\frac{x}{n}, y, \frac{z}{n}) \in E^*$, où n est un entier positif arbitraire et, en appliquant IV.3.3. $(n-1)$ fois, il vient :

$$\phi(x,y,z) = \phi(n \frac{x}{n}, y, n \frac{z}{n}) = n \phi(\frac{x}{n}, y, \frac{z}{n})$$

Comme $\phi(x,y,z)$ est fini, cela entraîne que $\lim_{n \rightarrow +\infty} \phi(\frac{x}{n}, y, \frac{z}{n}) = 0$.

Ce résultat montre que la fonction ϕ définie sur E^* peut se prolonger par continuité, sous certaines conditions, en $(0,y,0)$ et que la valeur prise se confond avec celle que nous avons fixée précédemment : $\phi(0,y,0) = 0$.

De la même manière, on démontrerait que l'application :

$$\psi(P(A), P(B), P(A.B)) = P(B) I(A/B)$$

est définie et continue sur E^* par ;

$$\psi(x,y,z) = y[F(x,y,z) - F(1,y,y)]$$

et que sur $E-E^*$ elle est définie par :

$$\psi(x,0,0) = 0 \quad \text{pour } x \in [0,1]$$

$$\psi(0,y,0) = +\infty \quad \text{pour } y \in]0,1]$$

Avec des notations évidentes, l'application ψ possède la propriété suivante :

Pour $(x,y_1,z_1) \in E$, $(x,y_2,z_2) \in E$ tels que $(x,y_1+y_2,z_1+z_2) \in E$, on a :

$$\psi(x,y_1+y_2,z_1+z_2) = \psi(x,y_1,z_1) + \psi(x,y_2,z_2)$$

En résumé, nous avons :

IV.3.4.- DEFINITION DE ϕ ET ψ .

Soient les applications ϕ et ψ définies sur E et à valeurs dans $\bar{\mathbb{R}}^+$ par :

$$\bullet \quad (x,y,z) \in E^* \quad \phi(x,y,z) = x[F(x,y,z) - F(x,1,x)]$$

$$\psi(x,y,z) = y[F(x,y,z) - F(1,y,y)]$$

$$\bullet \quad (x,y,z) \in E-E^*$$

$$\phi(0,y,0) = \psi(x,0,0) = 0 \quad \text{pour tout } (x,y) \in [0,1]^2$$

$$\phi(x,0,0) = \psi(0,y,0) = +\infty \quad \text{pour tout } (x,y) \in]0,1]^2$$

Si les hypothèses H.1 et H.2 sont vérifiées alors

l'application ϕ (resp. ψ) est finie et continue par rapport à chaque variable sur E_1 (resp. E_2) et vérifie la propriété suivante :

Pour $(x_1, y, z_1) \in E$, $(x_2, y, z_2) \in E$ tels que
 $(x_1+x_2, y, z_1+z_2) \in E$ on a :

$$\phi(x_1+x_2, y, z_1+z_2) = \phi(x_1, y, z_1) + \phi(x_2, y, z_2)$$

{Respectivement :

Pour $(x, y_1, z_1) \in E$, $(x, y_2, z_2) \in E$ tels que
 $(x, y_1+y_2, z_1+z_2) \in E$ on a :

$$\psi(x, y_1+y_2, z_1+z_2) = \psi(x, y_1, z_1) + \psi(x, y_2, z_2)$$

avec :

$$E = \{(x, y, z) \in [0, 1]^3 \mid \sup(0, x+y-1) \leq z \leq \inf(x, y)\}$$

$$E_1 = E \cap \{(x, y, z) \in [0, 1] \times]0, 1] \times [0, 1]\}$$

$$E_2 = E \cap \{(x, y, z) \in]0, 1] \times [0, 1]^2\}$$

$$E^* = E \cap \{(x, y, z) \in]0, 1]^2 \times [0, 1]\} = E_1 \cap E_2$$

Les solutions des équations fonctionnelles ainsi définies sont bien connues quand l'ensemble de définition est, par exemple, $\mathbb{R}^3$ ou $\mathbb{R}^{+3}$. L'ensemble de définition E de ϕ et ψ étant particulier, il convient de redémontrer rapidement ces résultats; E est convexe et il nous suffit d'adopter à ce cas les méthodes classiques de résolution [1].

LES EQUATIONS FONCTIONNELLES DE CAUCHY SUR E .

IV.3.5.- RAPPEL :

Equation fonctionnelle de Cauchy à une variable

Soit une application continue h de $[0, a]$ dans $\mathbb{R}$ (a strictement positif) qui vérifie, pour tout couple $(\alpha, \beta) \in [0, a]^2$ tel que $\alpha+\beta \in [0, a]$, la relation :

$$h(\alpha + \beta) = h(\alpha) + h(\beta)$$

L'unique solution de cette équation fonctionnelle est

$$h(\alpha) = C \alpha$$

où C est une constante quelconque; la constante C est positive si h est à valeur dans $\mathbb{R}^+$.

Utilisons la méthode classique. Pour $(\alpha_1, \alpha_2, \dots, \alpha_n) \in [0, a]^n$ tel que $\alpha_1 + \alpha_2 + \dots + \alpha_n \in [0, a]$ nous aurons :

$$h(\alpha_1 + \alpha_2 + \dots + \alpha_n) = h(\alpha_1) + h(\alpha_2) + \dots + h(\alpha_n)$$

car toutes les sommes partielles des α_i ($i = 1, \dots, n$) appartiennent à l'intervalle $[0, a]$. En particulier, si $\alpha_i = \alpha$ ($i = 1, \dots, n$) et $n \cdot \alpha \in [0, a]$ nous avons :

$$h(n\alpha) = n h(\alpha)$$

et si $\alpha_i = \frac{\alpha}{n}$ ($i = 1, \dots, n$) et $\alpha \in [0, a]$ nous avons :

$$h(\alpha) = h(n \frac{\alpha}{n}) = n h(\frac{\alpha}{n})$$

$$h(\frac{\alpha}{n}) = \frac{1}{n} h(\alpha)$$

Fixons-nous un $\alpha_0 \in]0, a]$ arbitraire, pour tous les éléments $\alpha \in [0, a]$ de la forme $\alpha = \frac{p}{q} \alpha_0$ avec p et q entiers positifs, nous avons donc :

$$h(\alpha) = h(\frac{p}{q} \alpha_0) = p h(\frac{\alpha_0}{q}) = \frac{p}{q} h(\alpha_0) = \alpha \cdot \frac{h(\alpha_0)}{\alpha_0}$$

Par continuité, pour les éléments $\alpha \in [0, a]$ de la forme $\alpha = \lambda \alpha_0$, où λ n'est pas un nombre rationnel, on obtient :

$$h(\alpha) = h(\lambda \alpha_0) = \lambda h(\alpha_0) = \alpha \frac{h(\alpha_0)}{\alpha_0}$$

La quantité $C = \frac{h(\alpha_0)}{\alpha_0}$ est une constante quelconque si h est à valeurs dans $\mathbb{R}$, c'est une constante positive si h est à valeur dans $\mathbb{R}^+$.

IV.3.6.- RAPPEL :

Equation fonctionnelle de Cauchy à deux variables.

Soit ϕ une application définie sur l'ensemble E_1 et à valeurs dans $\mathbb{R}^+$.

$$E_1 = \{(x,y,z) \in [0,1] \times]0,1] \times [0,1] \mid \text{Sup}(0,x+y-1) \leq z \leq \text{Inf}(x,y)\}$$

L'application ϕ est continue par rapport à chacune des variables x, z pour tout y fixé et vérifie la condition suivante :

Pour $(x_1, y, z_1) \in E_1$, $(x_2, y, z_2) \in E_1$ tels que $(x_1+x_2, y, z_1+z_2) \in E_1$ on a :

$$\phi(x_1+x_2, y, z_1+z_2) = \phi(x_1, y, z_1) + \phi(x_2, y, z_2)$$

L'unique solution de cette équation fonctionnelle est :

$$\phi(x,y,z) = K(y) x + H(y) z$$

où $K(y)$ et $H(y)$ sont des fonctions finies quelconques de y telles que

$$K(y) \geq 0 \quad \text{et} \quad H(y) \geq -K(y)$$

Notons, en premier lieu que $(x,y,z) \in E_1$ implique

- $(z,y,z) \in E_1$ car $(z,y,z) \in [0,1] \times]0,1] \times [0,1]$*
et

$$\text{Sup}(0,z+y-1) \leq z \leq \text{Inf}(y,z)$$

- $(x-z,y,0) \in E_1$ car $(x-z,y,0) \in [0,1] \times]0,1] \times [0,1]$*
et

$$\text{Sup}(0,x-z+y-1) \leq 0 \leq \text{Inf}(x-z,y)$$

Nous pouvons donc toujours écrire, pour $(x,y,z) \in E_1$:

$$\phi(x,y,z) = \phi(z,y,z) + \phi(x-z,y,0)$$

Fixons-nous y et faisons varier le couple (x,z) ; notre Problème revient à déterminer les solutions générales des deux équations fonctionnelles suivantes :

$$f_y(z) = \phi(z,y,z) \quad \text{avec } 0 \leq z \leq y.$$

$$f_y(z_1+z_2) = f_y(z_1) + f_y(z_2) \quad \text{pour } (z_1, z_2, z_1+z_2) \in [0,y]^3$$

$$g_y(t) = \phi(t,y,0) \quad \text{avec } 0 \leq t = x-z \leq 1-y$$

$$g_y(t_1+t_2) = g_y(t_1) + g_y(t_2) \quad \text{pour } (t_1, t_2, t_1+t_2) \in [0,1-y]^3$$

En utilisant les résultats du rappel IV.3.5., il vient :

$$f_y(z) = C(y).z \quad \text{pour } y > 0$$

$$g_y(t) = K(y).t \quad \text{pour } y < 1$$

où $C(y)$ et $K(y)$ sont des constantes positives qui ne dépendent que de y .

Nous obtenons donc, pour $0 < y < 1$:

$$\phi(x,y,z) = C(y) z + K(y) (x-z)$$

$$= K(y) x + H(y) z$$

avec $K(y) \geq 0$ et $H(y) \geq -K(y)$.

Pour $y=1$ on a $x=z$ et on obtient comme solution $\phi(x,1,x) = C x$. En posant $C = K(1) + H(1)$ on a donc

$$\phi(x,y,z) = K(y) x + H(y) z \quad \text{pour tout } y \in]0,1]$$

IV.3.7.- THEOREME.

Une condition nécessaire et suffisante pour qu'une information I définie sur l'espace mesuré $(\Omega, \mathcal{T}, P)$ vérifie les hypothèses 1 et 2 est :

Pour tous les couples $(A, B) \in \mathcal{A} \times \mathcal{B}$ on a :

$$I(A, B) = F(P(A), P(B), P(A \cap B))$$

L'application F est définie par :

$$F(x, y, z) = f_1(x) + f_2(y) - K\left(\frac{z}{xy} - 1\right) \quad \text{si } xy > 0$$

$$F(x, y, z) = +\infty \quad \text{si } xy = 0$$

dans cette formule :

a) K est une constante positive ou nulle quelconque.

b) f_1 et f_2 sont des applications non croissantes de $[0, 1]$ dans $\bar{\mathbb{R}}^+$, finies et continues sur $]0, 1]$ et telles que :

$$f_1(0) = f_2(0) = +\infty$$

$$f_1(1) = f_2(1) = 0$$

c) f_1 et f_2 vérifient la condition suivante :

Pour $(t, t') \in]0, 1]^2$ tel que $t < t'$ on a

$$f_i(t) - f_i(t') \geq K\left[\frac{1}{t} - \frac{1}{t'}\right] \quad i = 1, 2.$$

IV.3.8.- COROLLAIRE.

La décomposition de F ainsi définie est unique.

De plus pour $A \in \mathcal{A}$

$$I(A) = f_1(P(A)) = F(P(A), 1, P(A)) \geq K\left(\frac{1}{P(A)} - 1\right)$$

Pour $B \in \mathcal{B}$

$$I(B) = f_2(P(B)) = F(1, P(B), P(B)) \geq K\left(\frac{1}{P(B)} - 1\right)$$

Examinons en premier lieu les conséquences des hypothèses 1 et 2 que nous avons énoncées au paragraphe IV.3.4. et les résultats du rappel IV.3.6., sur les équations fonctionnelles de CAUCHY. Avec les notations du paragraphe IV.3.4., il est clair que :

$$\phi(x, y, z) = K(y) x + H(y) z \quad \text{pour } (x, y, z) \in E_1$$

en particulier, pour $x > 0$, c'est-à-dire pour $(x, y, z) \in E^*$ il vient :

$$x[F(x, y, z) - F(x, 1, x)] = K(y) x + H(y) z$$

ou

$$F(x, y, z) = K(y) + H(y) \frac{z}{x} + F(x, 1, x)$$

De même, pour $y > 0$, c'est-à-dire pour $(x, y, z) \in E^*$:

$$\psi(x, y, z) = K'(x) y + H'(y) z$$

et

$$F(x, y, z) = K'(x) + H'(x) \frac{z}{y} + F(1, y, y)$$

En identifiant les deux expressions de $F(x, y, z)$ pour $(x, y) \in]0, 1]^2$ fixé et en faisant varier z de telle manière que $(x, y, z) \in E^*$, il vient :

$$\cdot \quad \frac{H(y)}{x} - \frac{H'(x)}{y} = 0$$

$$\cdot \quad K(y) + F(x, 1, x) - K'(x) - F(1, y, y) = 0$$

On en déduit en faisant varier indépendamment x et $y \in]0, 1]^2$

$$\cdot \quad H(y) = \frac{C}{y} \quad \text{et} \quad H(x) = \frac{C}{x}$$

$$\cdot \quad K(y) = F(1, y, y) + K \quad \text{et} \quad K'(x) = F(x, 1, x) + K$$

où K et C sont des constantes indépendantes de $(x, y, z) \in E^*$.

$$F(x,y,z) = F(x,1,x) + F(1,y,y) + K + C \frac{z}{xy}$$

En remarquant que $I(\Omega) = F(1,1,1) = 0$, il vient $K = -C$ et comme, de plus, la proposition IV.3.6 implique $K(y) \geq 0$ c'est-à-dire $F(1,y,y) + K \geq 0$ pour tout $y \in [0,1]$, nous avons $K \geq 0$

$$F(x,y,z) = F(x,1,x) + F(1,y,y) + K(1 - \frac{z}{xy})$$

Il est clair, de plus, que, dans l'hypothèse H.2, tous les termes de cette formule sont finis pour $(x,y,z) \in E^*$ et que sur $E-E^*$

$$F(0,0,0) = F(0,y,0) = F(x,0,0) = +\infty \quad \text{pour tout } (x,y) \in [0,1]^2$$

Nous sommes donc dans les conditions prévues dans la proposition IV.2.8 en posant, pour $(x,y,z) \in E$:

$$f_1(x) = F(x,1,x)$$

$$f_2(y) = F(1,y,y)$$

$$\theta(x,y,z) = K(\frac{z}{xy} - 1) \quad \text{avec } K \geq 0 \quad \text{pour } (x,y,z) \in E^*$$

$$\theta(x,y,z) = 0 \quad \text{pour } (x,y,z) \in E-E^*$$

soit :

$$F(x,y,z) = f_1(x) + f_2(y) - \theta(x,y,z)$$

On déduit les conditions a et b du théorème des résultats précédents ainsi que de la proposition IV.2.8. Examinons les conséquences de la condition c de la proposition IV.2.8.

Pour $(x,y,z) \in E$ et $(x',y,z') \in E$ tels que $x'-x \geq z'-z \geq 0$ il faut :

$$f_1(x) - f_1(x') \geq \theta(x,y,z) - \theta(x',y,z')$$

c'est-à-dire, quand $(x,y,z) \in E^*$ et $(x',y,z') \in E^*$

$$f_1(x) - f_1(x') \geq K \left[\frac{z}{xy} - \frac{z'}{x'y} \right]$$

Fixons-nous le couple $(x, x') \in]0, 1]^2$ et faisons varier (y, z) et (y, z') de telle manière que $(x, y, z) \in E^*$, $(x', y, z') \in E^*$ et $x' - x \geq z' - z \geq 0$, on peut toujours obtenir (z, z', y) tel que $z = z' = y \leq x \leq x'$.

$$f_1(x) - f_1(x') \geq K \left[\frac{1}{x} - \frac{1}{x'} \right] = K \left[\frac{x' - x}{xx'} \right]$$

On obtient de même que, pour tout $(y, y') \in [0, 1]^2$ tels que $y' \geq y$, il faut

$$f_2(y) - f_2(y') \geq K \left[\frac{1}{y} - \frac{1}{y'} \right] = K \left[\frac{y' - y}{yy'} \right]$$

Nous avons ainsi démontré que les conditions sont nécessaires. On en déduit évidemment le corollaire IV.3.8. car :

$$f_1(t) - f_1(1) \geq K \left(\frac{1}{t} - 1 \right) \text{ et } f_1(1) = 0$$

Démontrons maintenant que les conditions sont suffisantes. Il est clair que l'hypothèse 1 doit être vérifiée puisque la formule IV.3.1. :

$$P(A_1 \cup A_2) I(B/A_1 \cup A_2) = P(A_1) I(B/A_1) + P(A_2) I(B/A_2)$$

est à la base de notre équation fonctionnelle. Faisons cependant rapidement le calcul, soit $B \in \mathcal{B}$ et (A_1, A_2) un couple d'ensembles disjoints de $\mathcal{A}$ ($A_1 \neq \emptyset$ et $A_2 \neq \emptyset$). Si $B = \emptyset$, on a $I(B) = I(B/A_1) = I(B/A_2) = +\infty$ et la relation est vérifiée. Si $B \neq \emptyset$, on a $P(B) > 0$, comme, de plus, $P(A_1) > 0$, $P(A_2) > 0$, les quantités $f_1(P(A_1))$ et $f_1(P(A_2))$ sont finies, et il vient

$$I(B/A_1) = f_2(P(B)) - K \left(\frac{P(A_1 \cdot B)}{P(A_1) P(B)} - 1 \right)$$

$$I(B/A_2) = f_2(P(B)) - K \left(\frac{P(A_2 \cdot B)}{P(A_2) P(B)} - 1 \right)$$

$$I(B/A_1 \cup A_2) = f_2(P(B)) - K \left(\frac{P((A_1 \cup A_2) \cdot B)}{P(A_1 \cup A_2) P(B)} - 1 \right)$$

Il est clair que

$$\begin{aligned} P(A_1 \cup A_2) I(B/A_1 \cup A_2) &= \\ &= P(A_1) f_2(P(B)) + P(A_2) f_2(P(B)) - K \frac{P(A_1 \cdot B)}{P(B)} - K \frac{P(A_2 \cdot B)}{P(B)} + K P(A_1) + K P(A_2) \end{aligned}$$

c'est-à-dire :

$$P(A_1 \cup A_2) I(B/A_1 \cup A_2) = P(A_1) I(B/A_1) + P(A_2) I(B/A_2)$$

et la proposition IV.2.3. nous permet d'affirmer que l'hypothèse 1 est bien vérifiée.

Montrons maintenant que l'hypothèse 2 est vérifiée, c'est-à-dire que l'on peut choisir arbitrairement K , f_1 , f_2 sous les seules conditions énoncées dans le théorème. Il est clair que toutes les conditions imposées dans la propositions IV.2.8. sont vérifiées sauf peut-être la condition c . Nous devons montrer que, pour tout $(x, y, z) \in E^*$, $(x', y, z') \in E^*$, tels que $x' - x \geq z' - z \geq 0$. On a

$$f_1(x) - f_1(x') \geq \theta(x, y, z) - \theta(x', y, z') = K \left[\frac{z}{xy} - \frac{z'}{x'y} \right]$$

Nous avons imposé que :

$$f_1(x) - f_1(x') \geq K \left[\frac{1}{x} - \frac{1}{x'} \right]$$

montrons que :

$$K \left[\frac{1}{x} - \frac{1}{x'} \right] \geq K \left[\frac{z}{xy} - \frac{z'}{x'y} \right]$$

c'est-à-dire

$$\frac{1}{x} \left[1 - \frac{z}{y} \right] \geq \frac{1}{x'} \left[1 - \frac{z'}{y} \right]$$

ce qui est évident car

$$x' - x \geq 0 \implies \frac{1}{x} \geq \frac{1}{x'}$$

$$z' - z \geq 0 \implies 1 - \frac{z}{y} \geq 1 - \frac{z'}{y} \geq 0$$

La formule se prolonge sans difficulté sur $E-E^*$ où nous avons posé $\theta(x,y,z) = 0$. On démontrerait de même la relation concernant l'application f_2 .

IV.3.9.- COROLLAIRE.

Une condition nécessaire et suffisante pour qu'une information I définie sur l'espace mesuré $(\Omega, \mathcal{I}, P)$ vérifie les hypothèses 1 et 2 en supposant de plus que l'application est dérivable sur E^ par rapport à chaque variable, est :*

Pour tous les couples $(A,B) \in \mathcal{A} \times \mathcal{B}$ on a

$$I(A,B) = F(P(A), P(B), P(A,B))$$

L'application F est définie par

$$F(x,y,z) = f_1(x) + f_2(y) - K\left(\frac{z}{xy} - 1\right) \quad \text{si } xy > 0$$

$$F(x,y,z) = +\infty \quad \text{sinon } xy = 0$$

dans cette formule.

- a) K est une constante positive ou nulle quelconque
- b) f_1 et f_2 sont des applications de $[0,1]$ dans $\bar{\mathbb{R}}^+$, finies, continues et dérivables sur $]0,1]$ et telles que :

$$\begin{aligned} f'_i(t) &\leq -\frac{K}{t^2} & i = 1,2 \\ f_i(0) &= +\infty & i = 1,2 \\ f_i(1) &= 0 & i = 1,2 \end{aligned}$$

Les conditions sont évidemment nécessaires ; elles sont aussi suffisantes car, de $f'_i(t) \leq -\frac{K}{t^2}$ on tire immédiatement, pour $t < t'$

$$\int_t^{t'} f'_1(u) du \leq \int_t^{t'} \left(-\frac{K}{u^2}\right) du$$

c'est-à-dire

$$f_1(t) - f_1(t') \geq K\left(\frac{1}{t} - \frac{1}{t'}\right)$$

En particulier $f_1(t) - f_1(1) = f_1(t) \geq K\left(\frac{1}{t} - 1\right)$

Posons, avec les notations du théorème IV.3.7. :

$$I_1(A.B) = \frac{1}{P(A)} - 1 + \frac{1}{P(B)} - 1 - \frac{P(A.B)}{P(A)P(B)} + 1$$

$$I_1(A.B) = \frac{P(A) + P(B) - P(A.B) - P(A)P(B)}{P(A)P(B)}$$

c'est-à-dire :

$$f_1(t) = f_2(t) = \frac{1}{t} - 1 \quad \text{et} \quad K = 1$$

$$F_1(x,y,z) = \frac{x+y-z-xy}{xy} = \left(\frac{1}{x} - 1\right) + \left(\frac{1}{y} - 1\right) - \left(\frac{z}{xy} - 1\right)$$

Il vient naturellement :

IV.3.10.- PROPOSITION.

Avec les notations du théorème IV.3.7., toutes les applications F qui vérifient les hypothèses 1 et 2 sont telles qu'il existe $K \geq 0$ tel que pour tout $K_0 \in [0, K]$ on a sur E :

$$F(x,y,z) \geq K_0 \cdot F_1(x,y,z)$$

K est la constante qui intervient dans la définition de F .

$$F(x,y,z) = f_1(x) + f_2(y) - K\left(\frac{z}{xy} - 1\right) \quad \text{si} \quad xy > 0$$

$$F(x,y,z) = +\infty \quad \text{si} \quad xy = 0$$

De plus :

$$f_1(x) \geq K\left(\frac{1}{x} - 1\right)$$

$$f_2(y) \geq K\left(\frac{1}{y} - 1\right).$$

IV.3.11.- COROLLAIRE.

Si K est strictement positif, les applications f_i , $i = 1, 2$, du théorème IV.3.7. sont toujours différentes de

$$c \cdot \text{Log} \frac{1}{x} \quad c > 0$$

C'est-à-dire que l'on ne peut pas obtenir l'information de Wiener-Shanon

$$I(A) = c \cdot \text{Log} \frac{1}{P(A)}$$

C'est une conséquence immédiate de la condition c du théorème IV.3.7.

IV.3.12.- PROLONGATION DE I SUR T .

Il est clair que l'information I n'est ainsi définie que pour les éléments de T de la forme $A.B$, $(A,B) \in A \times B$. En fait, ce sont les seuls éléments pour lesquels l'information nous intéresse. Montrons cependant qu'il existe différentes techniques permettant de prolonger I sur T .

Soit $C \in T$, définissons :

$$A_C = \cap \{A \in A \mid A \supset C\}.$$

$$B_C = \cap \{B \in B \mid B \supset C\}$$

$A_C \in A$ et $B_C \in B$ car nous travaillons sur des tribus comportant un nombre fini d'éléments.

Soit I une information vérifiant les hypothèses 1 et 2. Pour $(A,B) \in A \times B$ nous avons f_1, f_2, K tel que :

$$I(A,B) = f_1(P(A)) + f_2(P(B)) - K\left(\frac{P(A,B)}{P(A) \cdot P(B)} - 1\right)$$

$$= F(P(A), P(B), P(A,B))$$

On peut prolonger I sur T en prenant, par exemple :

$$I'(C) = f_1(P(A_C)) + f_2(P(B_C)) - K\left(\frac{P(C)}{P(A_C) \cdot P(B_C)} - 1\right)$$

$$I'(C) = F(P(A_C), P(B_C), P(C))$$

ou

$$I''(C) = f_1(P(A_C)) + f_2(P(B_C)) - K\left(\frac{P(A_C \cdot B_C)}{P(A_C) \cdot P(B_C)} - 1\right)$$

$$I''(C) = F(P(A_C), P(B_C), P(A_C \cdot B_C))$$

Il est clair que I' et I'' coïncident avec I sur les éléments A, B , $(A, B) \in A \times B$, que I' et I'' sont définis sur T tout entier, en particulier :

$$I'(C) \geq I''(C) = I(A_C \cdot B_C)$$

Il est trivial de montrer que I'' vérifie les axiomes A.1 et A.2 de I.1.1. : en particulier $C \subset C' \implies A_C \subset A_{C'}$, et $B_C \subset B_{C'}$, et il est facile d'en déduire, pour I'' :

$$I''(C) = I(A_C \cdot B_C) \geq I(A_C, B_{C'}) = I''(C') \geq 0$$

Pour I' la démonstration est un peu plus complexe :

$$f_1(P(A_C)) - f_1(P(A_{C'})) \geq K\left[\frac{1}{P(A_C)} - \frac{1}{P(A_{C'})}\right]$$

$$f_2(P(B_C)) - f_2(P(B_{C'})) \geq K\left[\frac{1}{P(B_C)} - \frac{1}{P(B_{C'})}\right]$$

pour montrer que $I'(C) \geq I'(C')$ il faut montrer que :

$$I'(C) - I'(C') \geq 0$$

or :

$$\begin{aligned}
I'(C) - I'(C') &\geq K\left[\frac{1}{P(A_C)} - \frac{1}{P(A_{C'})}\right] + K\left[\frac{1}{P(B_C)} - \frac{1}{P(B_{C'})}\right] \\
&\quad - K\left[\frac{P(C)}{P(A_C)P(B_C)} - 1\right] + K\left[\frac{P(C')}{P(A_{C'})P(B_{C'})} - 1\right] \\
I'(C) - I'(C') &\geq K\left[\frac{P(A_C)+P(B_C)-P(C)-P(A_C).P(B_C)}{P(A_C).P(B_C)} - \right. \\
&\quad \left. - \frac{P(A_{C'})+P(B_{C'})-P(C')-P(A_{C'}).P(B_{C'})}{P(A_{C'}).P(B_{C'})} \right]
\end{aligned}$$

C'est-à-dire qu'il faut montrer que la prolongation I est admissible pour la fonction F_1 de la proposition IV.3.10. :

$$F_1(x,y,z) = \frac{x+y-z-xy}{xy} = \frac{1}{x} + \frac{1}{y} - \frac{z}{xy} - 1$$

Cette fonction est continue et dérivable et toutes les dérivées partielles sont négatives sur E :

$$\frac{\partial F_1}{\partial x} = -\frac{1}{x^2} + \frac{z}{y} \frac{1}{x^2} = \frac{1}{x^2} \left(\frac{z}{y} - 1 \right) \leq 0$$

$$\frac{\partial F_1}{\partial y} = \frac{1}{y^2} \left(\frac{z}{x} - 1 \right) \leq 0$$

$$\frac{\partial F_1}{\partial z} = -\frac{1}{xy} \leq 0$$

On en déduit $\forall (x,y,z) \in E$ et $\forall (x',y',z') \in E$ tels que

$$x' \geq x, y' \geq y, z' \geq z$$

$$F_1(x,y,z) \geq F_1(x',y',z')$$

c'est-à-dire :

$$I(C) - I'(C') \geq K[F_1(P(A_C), P(B_C), P(C)) - F_1(P(A_{C'}), P(B_{C'}), P(C'))] \geq 0$$

On en déduit facilement que I' vérifie les axiomes 1 et 2 de I.1.1.

IV.4.- PROPRIETES.

Nous savons maintenant que toutes les informations I définies sur l'espace $(\Omega, \mathcal{T}, P)$ et vérifiant les hypothèses 1 et 2 sont de la forme, pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$:

$$I(A) = f_1(P(A))$$

$$I(B) = f_2(P(B))$$

$$I(A.B) = f_1(P(A)) + f_2(P(B)) - K \cdot \left(\frac{P(A.B)}{P(A).P(B)} - 1 \right)$$

la constante K et les applications f_1, f_2 sont définies dans le théorème IV.3.7.

Etant donné K strictement positif, il est évident que ;

$$I(A.B) = I(A) + I(B) \quad \Longleftrightarrow$$

$$P(A.B) = P(A) \cdot P(B)$$

même pour une information infinie ou une probabilité nulle.

IV.4.1.- PROPOSITION.

Etant donné une information I vérifiant les hypothèses 1 et 2, avec $K > 0$, l'indépendance en information des événements $A \in \mathcal{A}$ et $B \in \mathcal{B}$ implique leur indépendance en probabilité et inversement.

Etudions l'information conditionnelle apportée par $A \in \mathcal{A}$ connaissant l'information apportée par $B \in \mathcal{B}$.

$$I(A/B) = I(A.B) - I(B) = f_1(P(A)) - K\left(\frac{P(A.B)}{P(A).P(B)} - 1\right)$$

Soit $A \in \mathcal{A}$ fixé et $(B_1, B_2) \in \mathcal{B} \times \mathcal{B}$, il est clair que :

$$P(A/B_1) > P(A/B_2) \quad \Longleftrightarrow$$

$$I(A/B_1) < I(A/B_2)$$

IV.4.2.- PROPOSITION.

Soit une information I vérifiant les hypothèses 1 et 2, avec $K > 0$; étant donné $A \in \mathcal{A}$ fixé, l'information conditionnelle $I(A/B)$ apportée sur A par un événement $B \in \mathcal{B}$ est une fonction strictement décroissante de $P(A/B)$ et inversement.

En particulier :

$$P(A/B) = 0 \quad \Longleftrightarrow \quad I(A/B) = I(A) + K$$

$$P(A/B) = P(A) \quad \Longleftrightarrow \quad I(A/B) = I(A)$$

$$P(A/B) = 1 \quad \Longleftrightarrow \quad I(A/B) = I(A) - K\left(\frac{1}{P(A)} - 1\right)$$

de même en permutant les rôles de A et B .

Rappelons, IV.3.8, que :

$$I(A) \geq K\left(\frac{1}{P(A)} - 1\right)$$

Reprenons l'exemple du paragraphe IV.1.2., si Ω est un ensemble d'étudiants.

$$B_1 = \{\text{étudiants plein temps}\}$$

$$B_2 = \{\text{auditeurs salariés, étudiants à temps partiel}\}$$

$$A = \{\text{réussite à l'examen}\}$$

Il est naturel d'imaginer que si la probabilité $P(A/B_1)$ de réussir à l'examen pour un étudiant plein temps est plus grande que la probabilité $P(A/B_2)$ de réussir à l'examen pour un étudiant à temps partiel, l'information fournie en disant "cette personne a réussi l'examen" est plus importante quand il s'agit d'un étudiant à temps partiel que quand on parle d'un étudiant plein temps, et inversement.

Les deux relations que nous venons de mettre en relief nous semblent très importantes. D'une part on retrouve une des propriétés de base de l'information de Wiener-Shannon pour les couples $(A,B) \in A \times B$. En effet pour :

$$I(A,B) = C \cdot \log \frac{1}{P(A,B)} \quad C > 0$$

nous avons aussi la propriété :

$$I(A,B) = I(A) + I(B) \iff P(A,B) = P(A) \cdot P(B)$$

De plus

$$I(A/B) = C \log \frac{1}{P(A/B)}$$

est aussi une fonction décroissante de $P(A/B)$. Cependant l'information de Wiener-Shannon ne vérifie pas l'hypothèse 1, ce qui provient du fait que par hypothèse nous avons supposé

$$I(A,B) = F(P(A), P(B), P(A,B))$$

alors que dans le cadre de l'information de Wiener-Shannon on suppose uniquement

$$I(A,B) = F(P(A,B))$$

D'autre part il semble que l'on pourrait prendre les propositions IV.4.1. et IV.4.2. comme hypothèses de départ de ce chapitre à la place de notre hypothèse 2.

Notons que cette information nous donne un exemple d'information où un événement de mesure nulle a une information finie, c'est-à-dire qu'elle nous permet de comparer des événements de mesure nulle.

Si $P(A.B) = 0$ alors

$$I(A.B) = f_1(P(A)) + f_2(P(B)) + K$$

par exemple.

$$I(A.B) = \left\{ \frac{1}{P(A)} - 1 + \frac{1}{P(B)} \right\} \cdot K$$

avec

$$P(A) > 0 \quad \text{et} \quad P(B) > 0$$

Il est clair enfin que les propriétés de f_1 et f_2 définies dans le théorème IV.3.7. impliquent que les restrictions de l'information I à (Ω, A) et (Ω, B) sont composables (I.1.1 - Axiome 4) ; par contre, l'information I sur (Ω, T) n'est pas composable en général.

On vérifiera que pour l'information I_1 :

$$\forall (A, B) \in A \times B$$

$$I_1(A \cap B) = \frac{1}{P(A)} + \frac{1}{P(B)} - \frac{P(A \cap B)}{P(A)P(B)} - 1$$

l'application de la loi de composition F (I.1.1. - Axiome 4) à

$$\forall (A, B) \in A \times B \quad I_1(A) = F(I_1(A \cap B), I_1(A \cap B^c))$$

conduit à des contradictions.

Par exemple pour

$$P(A_1) = \frac{1}{3}, \quad P(B_1) = \frac{1}{3}, \quad P(A_1 \cap B_1) = \frac{1}{9}$$

$$\text{et} \quad P(A_2) = \frac{4}{13}, \quad P(B_2) = \frac{1}{2}, \quad P(A_2 \cap B_2) = \frac{1}{26}.$$

On obtient

$$I_1(A_1 \cap B_1) = I_1(A_2 \cap B_2) = 4$$

$$I_1(A_1 \cap B_1^c) = I_1(A_2 \cap B_2^c) = 2,5$$

$$\text{et } I_1(A_1) = 2 \neq I_1(A_2) = 2,25.$$

CHAPITRE V

CONCLUSIONS

DE LA PREMIERE PARTIE

Au terme de cette partie, il nous est plus facile de situer notre étude dans l'ensemble des travaux consacrés à la mesure de *l'information apportée sur un événement par un groupe d'observateurs*. Certains de nos résultats sont liés à des problèmes de fond concernant le développement de la théorie généralisée de l'information ; aussi élargirons nous notre réflexion en replaçant ce travail et nos hypothèses dans le contexte plus global qui a été proposé lors des "Rencontres en Théories de l'Information" de juin 1973. [19] [24].

Par ailleurs dans la suite de ce chapitre nous essaierons de cerner les directions de travail de la théorie de l'information sur un événement insuffisamment développées et qui nous paraissent les plus fructueuses pour les applications envisagées ; en particulier nous traitons en détail un exemple inspiré de celui de J. KAMPE DE FERIET dans [19], qui met en évidence l'importance de certains résultats précédents et leurs prolongements possibles.

V.1. - ELEMENTS DE COMPARAISON DE DIFFERENTS TRAVAUX SUR LES MESURES DE L'INFORMATION PAR UN ENSEMBLE D'OBSERVATEURS.

L'idée de lier l'information apportée sur la réalisation d'un événement aux observateurs de cet événement est apparue très rapidement dans le développement

de la théorie généralisée de l'information. Deux ans après avoir introduit en collaboration avec B. FORTE une axiomatique de l'information fournie par un événement, J. KAMPE DE FERIET propose les définitions suivantes [14] [15] :

Etant donné

- un ensemble Ω dont les points ω représentent des événements élémentaires ;

- un ensemble $\mathcal{O}$ dont les points ξ représentent des observateurs ; nous supposons que tout observateur ξ fait correspondre à tout événement d'une classe $C \subset P(\Omega)$ un nombre $J(A|\xi)$ vérifiant les axiomes postulés pour la mesure de l'information : $J(A|\xi)$ est une application de $C \times \mathcal{O}$ dans $\bar{\mathbb{R}}^+$;

$$J(\Omega|\xi) = 0, \quad J(\emptyset|\xi) = +\infty ;$$

$$B \subset A \implies J(B|\xi) \geq J(A|\xi) .$$

Il définit ensuite une mesure de probabilité λ sur l'espace mesurable des observateurs $(\mathcal{O}, F)$ et en imposant à J des conditions convenables, il énonce le résultat :

$$\text{la moyenne} \quad \bar{J}(A) = \int_{\mathcal{O}} J(A|\xi) . d\lambda$$

définit une mesure d'information sur (Ω, C) .

J. KAMPE DE FERIET reprend le problème d'une manière un peu différente dans [17] : les hypothèses de départ qu'il formule alors et que nous avons citées dans l'introduction de ce mémoire, l'amènent naturellement à travailler dans l'espace produit $(\hat{\Omega}, \hat{S})$: à chaque observateur $\xi \in \mathcal{O}$ est associé un espace mesurable (Ω_ξ, S_ξ) ; l'information du super-observateur H.Q. qui perçoit la totalité des phénomènes correspond pour nous à l'information apportée sur un événement par la totalité des observateurs de $\mathcal{O}$. L'étude porte à nouveau sur l'idée que des observateurs indépendants apportent globalement une information qui est la moyenne des informations apportées par chacun d'eux.

B. SCHWEIZER et A. SKLAR [30] [31] ont développé des travaux voisins concernant les mesures aléatoires d'information et les mesures de l'information par un ensemble d'observateurs.

• L'idée essentielle que nous avons mise en oeuvre dans le chapitre I est de moduler l'information apportée sur la réalisation d'un événement en fonction de l'importance des observateurs de cet événement. Les définitions ainsi introduites nous ont permis de préciser les différentes notions d'indépendances qui sont sous-jacentes aux travaux précédemment cités. Les résultats des paragraphes II.1, III.1 et III.3 qui recourent certains résultats des travaux cités, apparaissent comme le fruit d'une hypothèse particulière sur l'indépendance des observateurs, d'autres hypothèses pouvant être envisagées en relation avec le modèle concret que l'on cherche à définir. Ce sont certaines de ces possibilités que nous avons développées dans les chapitres précédents et que nous évoquerons à la fin de ce chapitre.

On trouve là une des différences essentielles entre notre travail et les études précédemment citées ; les constructions que nous proposons sont basées essentiellement sur des hypothèses concernant la dépendance ou l'indépendance des observateurs, alors que les travaux antérieurs portent presque exclusivement sur l'étude de la composabilité des informations. (I.1.1. - Axiome 4).

Non seulement aucune hypothèse de ce type n'apparaît dans notre travail, mais *les informations que nous obtenons ne sont pas composables en général.*

C'est B. FORTE qui a introduit l'idée de moduler l'information en fonction des observateurs [8], mais après avoir défini la notion de collecteur, il a lui aussi orienté son travail vers la composabilité. Un *collecteur* est une application ψ convenablement définie de manière que pour tout $A \in S$ et $(F_1, F_2) \in F \times F$ avec $F_1 \cap F_2 = \emptyset$, on ait avec nos notations :

$$J(A, (F_1 \cup F_2)) = \psi(J(A, F_1), J(A, F_2), \lambda(F_1), \lambda(F_2))$$

Le collecteur généralise donc la notion de moyenne pour le calcul de l'information.

V.2. - PERSPECTIVES GENERALES.

Lors des "Rencontres en Théories de l'Information" de juin 1973 il a été mis en relief [19] [24] que, sur un plan très général, on peut considérer que toute information est une application monotone d'un espace partiellement ordonné E dans un espace totalement (ou partiellement) ordonné F qui conserve l'ordre de tous les couples ordonnés de E et qui permet en plus de comparer des couples non ordonnés de E (I.1.2.).

Obtenir de l'information sur un espace partiellement ordonné E , c'est "ajouter de l'ordre" dans E .

Cette idée, très fructueuse, n'est pas nouvelle. Elle est insuffisante, semble-t-il, pour permettre un développement harmonieux de la théorie. Il faut donc ajouter une hypothèse supplémentaire. Actuellement, dans le cadre de la théorie généralisée de l'information, c'est la notion de composabilité qui est à la base des travaux les plus importants. Nous avons pensé qu'à côté de cette branche en pleine expansion, il serait bon de développer les notions d'indépendance. Notons d'ailleurs que, dans la forme générale que nous avons proposée dans [24], les notions d'indépendance et de composabilité apparaissent en dualité selon que l'on considère l'ordre ou l'ordre converse.

Pourquoi choisir de construire des informations à partir d'hypothèses d'indépendance plutôt que d'hypothèses de composabilité ?

Peut-être parce que le hasard nous a d'abord conduit à étudier une information, celle du chapitre IV, qui possède des propriétés intéressantes au niveau des applications et qui n'est pas composable ; mais aussi à cause d'un choix délibéré, nous pensons que les notions de dépendance et d'indépendance sont très primitives : imaginons deux expériences séparées :

- lancer un dé
- jeter une pièce de monnaie

intuitivement il est facile de comprendre que la probabilité d'obtenir en même temps "6" au dé et "pile" à la pièce, est égale au produit des probabilités d'obtenir "6" et d'obtenir "pile". Pour nous il apparaît aussi clairement que l'information apportée par la réalisation simultanée des événements "6" et "pile" doit être égale à la somme des informations apportées par la réalisation de "6" et la réalisation de "pile". *Tout au moins ces événements qui sont évidemment indépendants en structure doivent l'être également en information quelle que soit la manière dont on traduit cette indépendance dans le système d'axiomes et de définitions utilisé.*

V.3. - PERSPECTIVES PARTICULIERES.

Voyons d'abord les prolongements et les généralisations des résultats précédents qui nous paraissent les plus intéressants à développer, avant de proposer quelques directions originales de travail.

V.3.1. - LES DEFINITIONS DU CHAPITRE I.

Les définitions du chapitre I semblent adaptées aux buts que nous nous

sommes fixé. Notons cependant que nous n'avons pas réussi à résoudre d'une manière vraiment satisfaisante, à notre sens, la difficulté suivante : les informations conditionnelles $J(./(A,F))$, si elles vérifient les hypothèses de la définition 1.2.2, ne vérifient pas celle de 1.2.6. C'est-à-dire, en toute rigueur, qu'une information conditionnelle n'est pas vraiment une information. Il est donc possible que l'on soit amené, par la suite, à modifier la définition 1.3.4, de l'information conditionnelle, ou plus probablement la définition 1.2.6. des valeurs universelles de l'information et de l'ensemble F_0 des groupes d'observateurs qui n'apportent pas d'information. En particulier, imposer aux éléments $F_0 \in F_0$ qui vérifient déjà $[A \in S \implies J(A, F_0) = 0]$ de vérifier aussi :

$$\forall (A, F) \in S \times F \implies J(A, (F \cup F_0)) = J(A, F)$$

permet de résoudre partiellement le problème posé.

Si une information conditionnelle $J(./(A,F))$ était une véritable information, on pourrait sans difficulté définir les notions d'événements ou d'observateurs conditionnellement indépendants, la notion d'information conditionnelle d'une information conditionnelle, etc... Les hypothèses du type de celle du paragraphe III 2 ou celles que nous allons évoquer ci-après, pourraient être alors formulées d'une manière à la fois plus élégante et plus rigoureuse.

V.3.2. - LES RESULTATS DES CHAPITRES II et III.

Nous avons signalé dans le paragraphe III.3 le problème de l'existence d'une décomposition régulière de J dans le cas d'un ensemble O quelconque.

L'étude du paragraphe III.2, dans lequel nous divisons l'ensemble fini O en deux ensembles disjoints O_1 et O_2 , devrait pouvoir sans trop de difficul-

tés être généralisée à un ensemble O quelconque ; peut-être, en prenant le problème d'une manière différente et en supposant que O est un ensemble produit. On peut aussi imaginer de reprendre les hypothèses de ce paragraphe en partitionnant O en trois (ou plus) ensembles disjoints O'_1, O'_2, O'_3 .

V.3.3. - LES RESULTATS DU CHAPITRE IV.

Deux généralisations viennent immédiatement à l'esprit : d'une part supposer que les tribus A et B sont quelconques ; d'autre part, en liaison avec les remarques précédentes (V.3.2.), étudier le cas de 3 observateurs, ou plus.

Nous pensons qu'avant de se lancer dans ce travail, il conviendra de bien réfléchir sur le sens de l'hypothèse 1 du chapitre IV :

$$\forall A \in \mathcal{A} \quad I(A) = \sum_{B \in \pi_B} P(B) \cdot I(A/B)$$

et de comparer cette formule à celle que nous avons obtenue au paragraphe III.1

$$\forall A \in \mathcal{S} \quad I(A) = \sum_{\xi \in \mathcal{O}} \lambda(\xi) \cdot I_{\xi}(A)$$

Au-delà de la similitude formelle des deux relations, il apparaît clairement que l'hypothèse 1 du chapitre IV fait jouer aux événements $B \in \pi_B$ le rôle d'observateurs pour les événements $A \in \mathcal{A}$. Sous-jacente à cette remarque et liée au problème de la définition de l'information conditionnelle (V.3.1) il y a l'idée que, *dans certains cas, une catégorie d'ensembles peut jouer alternativement le rôle d'événements ou d'observateurs pour une autre catégorie d'ensembles.*

V.4. - EXEMPLE.

L'étude d'un exemple présenté par J. KAMPE DE FERIET dans [19] va nous permettre de préciser nos idées sur une manière d'utiliser les résultats de notre travail et ouvrir une nouvelle perspective de développement dans l'optique proposée dans cette citation de R. THOM ([37], p. 19) :

"Tout modèle comporte a priori deux parties : une cinématique, dont l'objet est de paramétrer les formes ou les états du processus considéré ; une dynamique dont l'objet est de décrire l'évolution temporelle entre ces formes".

V.4.1. - A PROPOS DE L'EXEMPLE DES "HOMMES POLITIQUES".

"Considérons les n hommes politiques : $h_1, \dots, h_n$, dont nous avons toutes raisons de croire que, en vertu de leurs pouvoirs, ils déterminent par leurs décisions le proche avenir de l'histoire. A notre point de vue, chacun d'eux ne peut-être que dans deux états ; l'espace des phases Ω_j , représentant les états de h_j comprend donc deux points ; si h_j est au pouvoir nous poserons $\omega_j = 1$, s'il n'est plus au pouvoir (mort, déchu, non-réélu, etc...) $\omega_j = 0$: $\Omega_j = II = \{0, 1\}$.

Admettons que l'état d'un homme politique est indépendant (*) de l'état de tous les autres, c'est-à-dire que l'état de tous les autres étant fixé, il peut être lui-même soit dans l'état 1 soit dans l'état 0 ; l'espace des phases Ω représentant, à un instant donné, l'état des n hommes politiques est alors l'espace produit :

$$\Omega = \Omega_1 \times \dots \times \Omega_n = II^n ;$$

le point ω représentant un des 2^n états possibles est alors

$$\omega = (\omega_1, \dots, \omega_n) , \omega_j \in II , \forall j .$$

A l'instant initial tous les hommes politiques sont au pouvoir : $\omega(0) = (1, \dots, 1)$.
Aussi longtemps qu'il ne s'est produit aucun événement (disparition d'un homme politique) : $\omega(t) \equiv \omega(0)$, nous n'avons reçu aucune information et nous posons :

$$J[\omega(t)] = 0 .$$

Si à l'instant t_0 nous apprenons qu'un homme politique h_j est passé de l'état 1 à l'état 0, nous posons : $J[\omega(t)] = \beta_j \quad \beta_j \in \overline{R^+} \quad t \geq t_0$.

En raison de l'indépendance (**) des états des hommes politiques, nous aurons à un instant t quelconque :

$$J[\omega(t)] = \sum_1^n (1-\omega_j) \beta_j .$$

Notons qu'un événement est défini ici, non pas comme un état, c'est-à-dire un point de Ω , mais comme le passage d'un point de Ω à un autre."

Remarquons d'abord que le terme d'indépendance est utilisé deux fois.
Dans le premier cas (*) nous pensons que l'indépendance dont il s'agit est sans rapport direct avec l'information, mais est liée à la structure du modèle proposé ; c'est une hypothèse du type M-indépendance. C'est le deuxième cas (**)

qui nous intéresse du point de vue de l'information et nous allons essayer de retrouver la formule obtenue en mettant en évidence les différentes articulations du raisonnement et les hypothèses sous-jacentes.

a) INFORMATION STATIQUE : (Ω, A, I)

Admettons l'hypothèse (*); l'ensemble Ω a donc la structure d'ensemble produit :

$$\Omega = \prod_{j=1}^n \Omega_j$$

Soit $A_j = P(\Omega_j) = \{\Omega_j, \{1\}, \{0\}, \emptyset\}$; notons A l'algèbre de Boole produit des $\{A_j | j = 1, \dots, n\}$:

$$A = \bigotimes_{j=1}^n A_j$$

Il y a isomorphisme entre A_j et la sous-algèbre de A des cylindres de la forme :

$$A_j \times \prod_{\substack{i=1 \\ i \neq j}}^n \Omega_i, \quad A_j \in A_j$$

que nous identifierons au niveau des notations.

En dehors de toute idée de variation de l'état du système en fonction du temps, essayons de définir à un instant fixé une information I au sens I.1.1. sur (Ω, A) . Nous voulons attacher à I le sens suivant : Information apportée en apprenant que le système est dans l'état $\omega \in \Omega$.

Donnons un premier sens précis à l'hypothèse (**): *les sous-tribus* $\{A_j | j = 1, \dots, n\}$ de A sont *I-indépendantes en information*.

Les résultats des paragraphes II.1 et III.1 impliquent :

$$\forall A \in \mathcal{A}, I(A) = \sum_{j=1}^n I(A_j)$$

ou A_j est la projection de A sur Ω_j .

(Notons que, pour obtenir rigoureusement ce résultat, on peut utiliser le théorème II.1.4 et la proposition III.1.3, c'est-à-dire introduire un ensemble $O = \{1, \dots, n\}$ d'observateurs, chaque observateur étant lié à un homme politique, et il faut, en tout cas, supposer que, pour tout groupe F_1 et F_2 d'observateurs de O disjoints, les éléments de $\bigotimes_{i \in F_1} A_i$ et $\bigotimes_{i \in F_2} A_i$ sont *I-indépendants*).

En particulier, pour tout $\omega \in \Omega$, on a :

$$I(\omega) = \sum_{j=1}^n I(\omega_j)$$

en assimilant par abus de notation :

$$\omega \in \Omega \quad \text{à} \quad \{\omega\} \in A$$

et

$$\omega_j \in \Omega_j \quad \text{à} \quad \{\omega_j\} \times \prod_{\substack{i=1 \\ i \neq j}}^n \Omega_i \in A$$

Ce que l'on peut noter aussi :

$$I(\omega) = \sum_{j=1}^n (\omega_j \cdot \alpha_j + (1-\omega_j) \cdot \beta_j)$$

où l'on a $I(\omega_j) = \alpha_j \in \bar{\mathbb{R}}^+$ si $\omega_j = 1$

$I(\omega_j) = \beta_j \in \bar{\mathbb{R}}^+$ si $\omega_j = 0$

En particulier, avec les notations de l'exemple :

("A l'instant t_0 nous apprenons qu'un homme politique h_j est passé de l'état 1 à l'état 0")

$$\omega(0) = (1, \dots, 1, 1_j, 1, \dots, 1)$$

$$\omega(t_0) = (1, \dots, 1, 0_j, 1, \dots, 1)$$

Nous notons $\omega(0) \cup \omega(t_0)$ l'événement

$$\omega(0) \cup \omega(t_0) = \{(1, \dots, 1, 1_j, 1, \dots, 1), (1, \dots, 1, 0_j, 1, \dots, 1)\}$$

$$= \{1\} \times \dots \times \{1\} \times \Omega_j \times \{1\} \times \dots \times \{1\}$$

$$I(\omega(0)) = \sum_{i=1}^n \alpha_i$$

$$I(\omega(t_0)) = \sum_{\substack{i=1 \\ i \neq j}}^n \alpha_i + \beta_j$$

$$I(\omega(0) \cup \omega(t_0)) = \sum_{\substack{i=1 \\ i \neq j}}^n \alpha_i$$

Le triplet (Ω, A, I) ainsi défini nous donne donc l'information apportée par la connaissance de l'état $\omega \in \Omega$ du système, en dehors de tout contexte, a priori pourrait-on dire. Essayons maintenant de définir l'information apportée par le passage d'un état $\omega(s)$ à l'instant s à l'état $\omega(t)$ à l'instant t .

b) INFORMATION DE TRANSITION : $(\Omega^{(s)} \times \Omega^{(t)}, A^{(s)} \otimes A^{(t)}, J)$

Etant donné (Ω, A, I) nous pouvons associer à chaque $A \in \mathcal{A}$ son information $I(A)$; le système que l'on étudie évolue dans le temps ; *comment quantifier le gain d'information apporté par le passage du système de l'état A à l'instant s à l'état B à l'instant $t = s+1$?*

Associons à ces deux instants une réplique de l'espace (Ω, A)

$$(\Omega^{(s)}, A^{(s)}) = (\Omega^{(t)}, A^{(t)}) = (\Omega, A)$$

et définissons sur les cylindres de $(\Omega^{(s)} \times \Omega^{(t)}, A^{(s)} \otimes A^{(t)})$ une application J , définie en fonction de I sur (Ω, A) par :

$$\forall (A, B) \in A^{(s)} \times A^{(t)} \quad J(A \times B) = I(A) + I(B/(A \cup B))$$

d'après I.1.1.b. il vient $J(A \times B) = I(A) + I(B) - I(A \cup B)$

Supposons que J est une information au sens de J. KAMPE DE FERIET (I.1.1.) (ce qui n'est pas nécessairement vrai); nous aurons, pour tout $A \in A^{(s)}$ fixé :

$$B = A \Rightarrow J(A \times B) = I(A)$$

$$B = A^c \Rightarrow J(A \times B) = I(A) + I(B)$$

comme $J(A \times \Omega^{(t)}) = I(A)$ on a aussi

$$B = A \Rightarrow J((A \times B)/(A \times \Omega^{(t)})) = 0$$

$$B = A^c \Rightarrow J((A \times B)/(A \times \Omega^{(t)})) = I(B)$$

et, d'une manière générale :

$$0 \leq J((A \times B)/(A \times \Omega^{(t)})) = I(B/(A \cup B)) \leq I(B)$$

L'information conditionnelle $J((A \times B)/(A \times \Omega^{(t)}))$ semble pouvoir s'interpréter comme l'information apportée par le passage du système de l'état A , à l'instant s , à l'état B , à l'instant $t = s+1$.

Reprenons l'exemple étudié avec les valeurs de l'information I définie dans le paragraphe a

$$\begin{aligned} J(\omega(0) \times \omega(t_0)) &= I(\omega(0)) + I(\omega(t_0)) - I(\omega(0) \cup \omega(t_0)) \\ &= I(\omega(0)) + \beta_j \end{aligned}$$

$$J((\omega(0) \times \omega(t_0))/(\omega(0) \times \Omega^{(t_0)})) = \beta_j$$

A l'instant t_0 nous avons appris le passage du système de l'état $\omega(0)$ à l'état $\omega(t_0)$, le gain d'information correspondant est égal à β_j .

c) INFORMATION DYNAMIQUE : $(\Omega^T \times T, A^T \times F)$

Il nous reste à retrouver la dernière formule qui exprime l'augmentation de l'information au fur et à mesure du temps.

Notons T l'intervalle $[0, T_0] \subseteq \mathbb{N}$ qui représente le temps (discret), $F = P(T)$. Supposons que $\forall t \in T : (\Omega^{(t)}, A^{(t)}) = (\Omega, A)$ nous notons :

$$\begin{aligned}\Omega^T &= \Omega^{(0)} \times \dots \times \Omega^{(t)} \times \dots \times \Omega^{(T_0)} = \prod_T \Omega^{(t)} \\ A^T &= A^{(0)} \otimes \dots \otimes A^{(t)} \otimes \dots \otimes A^{(T_0)} = \otimes_T A^{(t)}\end{aligned}$$

la trajectoire $(\omega(0), \omega(1), \dots, \omega(t), \dots, \omega(T_0))$ définit les états successifs du système.

Soit J une information au sens I.2.2., I.2.6. sur $(\Omega^T \times T, A^T \times F)$ où le temps T joue le rôle de l'ensemble des observateurs, J vérifiant les conditions du théorème II.1.4. Explicitons une hypothèse du type de celle que l'on fait dans les processus de MARKOV :

Pour tout $A \in A^T$, l'information apportée sur A à l'instant t , connaissant l'information apportée sur A pendant l'intervalle de temps $[0, t-1]$ est égale à l'information apportée sur A à l'instant t connaissant l'information apportée sur A à l'instant $t-1$, en assimilant par abus de notation $t \in T$ à $\{t\} \in F = P(T)$:

$$J((A, t)/(A, [0, t-1])) = J((A, t)/(A, \{t-1\}))$$

On obtient par itérations successives :

$$J(A, [0, t]) = J(A, 0) + \sum_{t'=1}^t J((A, t')/(A, \{t'-1\}))$$

Nous avons supposé, par hypothèse, que J vérifie les conditions du théorème II.1.4 ; c'est-à-dire qu'on peut définir sur (Ω^T, A^T) une information I_T au sens de J. KAMPE DE FERIET, telle que :

$$\forall A \in A^T, I_T(A) = J(A, [0, T_0]) = J(A, T)$$

Notons $A^0, A^1, \dots, A^t, \dots, A^T$ les projections de $A \in A^T$ respectivement sur $\Omega^{(0)}, \Omega^{(1)}, \dots, \Omega^{(t)}, \dots, \Omega^{(T)}$; on a aussi :

$$J(A, 0) = I_T(A^0)$$

$$J((A, t)/(A, (t-1))) = I_T(A^t/A^{t-1})$$

Supposons que, sur chaque espace

$$(\Omega^{(t-1)} \times \Omega^{(t)}, A^{(t-1)} \otimes A^{(t)})$$

on ait défini l'information J du paragraphe b, et que

$$I_T(A^{t-1} \times A^t) = J(A^{t-1} \times A^t)$$

on aura :

$$I_T(A^t/A^{t-1}) = J((A^{t-1} \times A^t)/(A^{t-1} \times \Omega^{(t)}))$$

Il vient, en remplaçant J par sa valeur en fonction de l'information I sur (Ω, A) du paragraphe a

$$I_T(A^t/A^{t-1}) = I(A^t/(A^t \cup A^{t-1}))$$

$$J(A, [0, t]) = I(A_0) + \sum_{t'=1}^t I(A^{t'}/(A^{t'} \cup A^{t'-1}))$$

ce qui donne, pour l'exemple étudié (*en supposant* que $\omega(0) = (1, \dots, 1)$ et que les seuls changements possibles dans le temps sont des passages de l'état 1 à l'état 0),

$$J(A, [0, t]) = I(\omega(0)) + \sum_{j=1}^n (1 - \omega_j(t)) \cdot \beta_j$$

ou encore :

$$I_T(\omega(0), \omega(1), \dots, \omega(t)/\omega(0)) = \sum_{j=1}^n (1 - \omega_j(t)) \cdot \beta_j$$

c'est-à-dire le résultat proposé dans l'exemple cité.

V.4.2.- CONCLUSIONS DE L'EXEMPLE.

Dans les chapitres précédents, nous ne faisons aucune hypothèse concernant une éventuelle structure de l'ensemble $\mathcal{O}$ des observateurs. L'étude de l'exemple ci-dessus et en particulier le paragraphe V.4.1.c, où nous avons fait jouer au temps T le rôle de l'ensemble des observateurs, nous conduit naturellement à imaginer qu'il peut exister sur $\mathcal{O}$ une structure particulière qui implique a priori des relations entre les observateurs. Une relation d'ordre totale nous amènerait par exemple à introduire des hypothèses comparables à celles que l'on fait dans l'étude de certains types de processus stochastiques ou, si l'on suppose que $\mathcal{O}$ est un espace métrique, à introduire des hypothèses d'indépendance et de dépendance en fonction de la distance des observateurs ou des groupes d'observateurs, etc...

Notons aussi que le problème soulevé dans le paragraphe V.4.1.b est très intéressant : connaissant l'information *statique* que l'on a sur l'état ω d'un système (Ω, A) en déduire une information *dynamique* qui tient compte du gain d'information apporté par le passage d'un état à un autre; ce gain d'information devant être faible s'il y a peu de différence entre les deux états successifs, important sinon.

V.5. - INFORMATION ET PROBABILITE.

Dans [13] J. KAMPE DE FERIET explicite une des raisons qui l'ont conduit, en collaboration avec B. FORTE, à introduire une axiomatique de l'information indépendamment de la notion de probabilité :

"A notre point de vue l'attribution d'une probabilité $P(A)$ à un événement A est essentiellement liée à la possibilité de répéter indéfiniment cet événement ; si l'on désigne par $n(A)$ le nombre de réalisations de A dans une série de n épreuves indépendantes, la loi forte des grands nombres affirme que :

$$\lim_{n \rightarrow +\infty} \frac{n(A)}{n} = P(A)$$

la probabilité apparait ainsi a posteriori comme la limite de la fréquence de l'événement A .

La définition classique de l'information est alors équivalente à

$$(1) \quad I(A) = \lim_{n \rightarrow +\infty} \log \frac{n}{n(A)}$$

la valeur de l'information fournie par l'événement A apparaissent comme la limite du logarithme de l'inverse de la fréquence dans une série d'épreuves répétées indépendantes ; c'est à cette formule que sont dus, jusqu'ici, les succès de la théorie.

Mais en adoptant (1) nous nous interdisons de donner un sens à la mesure de l'information fournie par des événements, qui, par leur nature même, ne peuvent pas être répétés. Or il arrive constamment, dans notre vie quotidienne, que des événements A et B se produisent, dont la répétition dans une série d'épreuves

est inconcevable, et pour lesquels cependant nous affirmons en toute certitude :

$$I(A) < I(B) .$$

C'est pourquoi nous nous sommes proposés, en collaboration avec B. FORTE, d'énoncer un système d'axiomes pour $I(A)$ traduisant aussi simplement que possible l'idée que notre intuition se fait de l'information fournie par un événement, sans nous référer à la notion de probabilité".

Les résultats de la première partie de ce mémoire, exceptés éventuellement ceux du chapitre IV, conduisent effectivement à des mesures d'information "fournies par des événements, qui, par leur nature même, ne peuvent pas être répétés", et cela indépendamment de toute notion de probabilité ou de fréquence. C'est le cas en particulier pour l'exemple traité au paragraphe précédent.

Par contre, dans la deuxième partie de ce travail et pour les applications que nous envisageons, les notions de fréquence et de probabilité s'introduisent naturellement. Comme dans la théorie classique c'est la notion d'information moyenne fournie par les éléments d'une partition qui ouvre de larges perspectives d'application. C'est pourquoi nous avons été naturellement conduit à construire des informations à partir des résultats précédents mais en utilisant aussi les fréquences ainsi définies. Nous montrerons clairement dans le chapitre VI comment nos résultats englobent et généralisent les résultats de la théorie classique de l'information.

DEUXIEME PARTIE

INFORMATION SUR UNE PARTITION

Nous n'hésitons pas à reprendre, dans cette deuxième partie, les définitions et certains des résultats précédents en les explicitant de manière intuitive, afin que le lecteur qui s'intéresse surtout aux applications puisse aborder directement ces deux chapitres, après avoir parcouru le chapitre I. Nous cherchons essentiellement à présenter quelques perspectives d'applications, quitte à donner un peu moins d'importance à la rigueur mathématique formelle. Le lecteur peut toujours, le cas échéant, se référer aux énoncés de la première partie ou aux travaux cités pour préciser tel ou tel point des hypothèses ou des notations utilisées.

Tous les ensembles qui interviennent dans cette parties : Ω , $\hat{\Omega}$, $\emptyset$, A , B , S , $\hat{S}$, F sont finis. Les énoncés et les formules sont simplifiés en conséquence sans que nous rappelions à chaque fois cette convention.

CHAPITRE VI

INFORMATION MOYENNE.

La plupart des applications de la théorie classique de l'information sont basées sur la formule de SHANNON définissant l'information moyenne apportée par les éléments d'une partition et sur les propriétés de l'information mutuelle de deux partitions. Certaines des applications que nous envisageons utilisent elles aussi ces notions d'information apportée par une partition ; nous commençons donc par rappeler les propriétés qui ont été proposées pour définir axiomatiquement l'information apportée par une "*expérience*". L'information moyenne se présente alors comme un manière, parmi d'autres, de lier la notion d'information apportée par un événement à celle d'information apportée par une partition.

En associant de manière naturelle à chaque observateur la partition qu'il distingue nous pouvons définir l'information apportée par un observateur et surtout la notion d'information mutuelle apportée par un couple d'observateurs ou par deux groupes disjoints d'observateurs. Nous montrons en particulier que les résultats des chapitres précédents permettent de définir des informations mutuelles sur des couples de partitions dont les propriétés englobent et généralisent celles de l'information au sens de SHANNON.

Nous essaierons de montrer sur des exemples dans le chapitre suivant en quoi ces quantités nous semblent plus adaptées que les notions similaires de la théorie classique de l'information pour les applications envisagées.

VI.1. - INFORMATION APPORTEE PAR UNE PARTITION.

Etant donné un espace probabilisé (Ω, S, P) et Π une partition finie S -mesurable de Ω , nous supposons que le résultat d'une *expérience* est l'un des événements $A \in \Pi$ sans que l'on puisse dire lequel. *L'information moyenne qui est apportée par cette expérience, quand on la renouvelle un grand nombre de fois, est égale à*

$$H^*(\Pi) = \sum_{A \in \Pi} P(A) \cdot I^*(A) \quad (1)$$

ou $I^*(A)$ est l'information apportée par l'événement $A \in S$, par exemple

$$I^*(A) = C \cdot \log \frac{1}{P(A)} \quad C > 0$$

dans la formule de SHANNON.

Dès le début des travaux sur la théorie généralisée de l'information, deux directions d'études axiomatiques se sont dessinées : d'une part celle de l'information apportée par un événement [12], [13], d'autre part celle de l'information apportée par une expérience ou une partition. Notons en particulier les travaux relatifs à ce sujet de P. BENVENUTI, B. FORTE N. PINTACUDA [2] [9] [10] [11] dont nous nous inspirerons par la suite. La formule (1) n'est alors qu'une possibilité, parmi d'autres, de relier ces deux systèmes d'axiomes.

Notre but dans ce paragraphe n'est pas de faire une étude exhaustive de la notion d'information apportée par une partition ou une expérience, mais simplement, en nous inspirant des travaux cités et des idées que nous avons émises avec G. COMYN dans [6], [24], de définir avec précision la notion d'information apportée par une partition, de rappeler les propriétés

que nous aurons à utiliser et d'appliquer ces notions à des formules du type (1).

VI.1.1. - TREILLIS DES PARTITIONS.

Conformément aux définitions de I.1.2, commençons par définir l'espace et la relation d'ordre que nous utilisons. Supposons que Ω est un ensemble fini, non vide, quelconque, $S = P(\Omega)$. Une partition Π de Ω est un ensemble fini de parties de S , tel que :

- . $A \in \Pi \implies A \neq \emptyset$
- . $(A, A') \in \Pi \times \Pi \implies A = A' \text{ ou } A \cap A' = \emptyset$
- . $\bigcup_{A \in \Pi} A = \Omega$

Soit E l'ensemble des partitions de Ω . Une partition Π_1 est dite "plus fine" qu'une partition Π_2 si :

$$\forall A_1 \in \Pi_1, \exists A_2 \in \Pi_2 \text{ tel que } A_1 \subseteq A_2$$

et nous noterons :

$$\Pi_1 \leq \Pi_2.$$

Cette relation est un ordre partiel, reflexif sur E ; rappelons les propriétés de E qui nous intéressent [4]. Tout couple de partitions Π_1, Π_2 possède un infimum et un supremum. En particulier, nous noterons $\Pi_1 \wedge \Pi_2$ l'infimum de Π_1 et Π_2 :

$$\Pi_1 \wedge \Pi_2 = \{A \cap B \mid A \in \Pi_1, B \in \Pi_2\}.$$

De plus E possède un majorant universel $\Pi_M = \{\Omega\}$ et un minorant universel Π_m :

$$\Pi_m = \{\{\omega\} \mid \omega \in \Omega\}.$$

Ces propriétés suffisent pour pouvoir définir une information sur E . Rappelons cependant que, pour la relation d'ordre ainsi définie, E est un treillis géométrique (relativement complété, semi-modulaire et atomique). [4]

VI.1.2. - INFORMATION SUR E . (I.1.2. et les travaux cités dans l'introduction du paragraphe V.1)

• Une information sur l'ensemble E des partitions de Ω est une application H définie sur E et qui vérifie les axiomes suivants :

A.1) $H : E \rightarrow \bar{\mathbb{R}}^+$

A.2) L'information H est décroissante pour la relation d'ordre $\leq$ sur E :

$$(\Pi_1, \Pi_2) \in E \times E \text{ et } \Pi_1 \leq \Pi_2 \implies H(\Pi_1) \geq H(\Pi_2)$$

A.3) Deux partitions Π_1 et Π_2 sont dites indépendantes pour l'information H si :

$$H(\Pi_1 \cap \Pi_2) = H(\Pi_1) + H(\Pi_2).$$

Nous admettrons comme valeur universelle :

$$H(\{\Omega\}) = 0.$$

Il est naturel d'imaginer que l'information apportée par la partition formée du seul ensemble Ω est nulle ; cela implique aussi que cette partition $\{\Omega\}$ est indépendante de toutes les autres partitions. Par contre, nous ne ferons aucune hypothèse sur $H(\Pi_m)$; en particulier nous ne supposerons pas que $H(\Pi_m) = +\infty$; nous obtenons dans la suite des valeurs finies pour cette quantité et il n'y a, semble-t-il, aucune raison pour que la partition la plus fine Π_m

soit indépendante de toutes les autres partitions.

• L'information conditionnelle fournie par la partition Π_1 connaissant l'information apportée par la partition Π_2 est définie par :

$$H(\Pi_1/\Pi_2) = H(\Pi_1 \cap \Pi_2) - H(\Pi_2) \text{ si } H(\Pi_2) < +\infty.$$

En posant le cas échéant $H(\Pi_1/\Pi_2) = H(\Pi_1)$ si $H(\Pi_2) = +\infty$ on rend équivalentes, pour tout couple $(\Pi_1, \Pi_2) \in E \times E$, les propositions

" Π_1 est H-indépendante de Π_2 " et " $H(\Pi_1/\Pi_2) = H(\Pi_1)$ ".

• L'information mutuelle apportée par les partitions Π_1 et Π_2 est définie par :

$$H(\Pi_1; \Pi_2) = H(\Pi_1) + H(\Pi_2) - H(\Pi_1 \cap \Pi_2)$$

si $H(\Pi_1) < +\infty$ et $H(\Pi_2) < +\infty$.

La convention

$$H(\Pi_1; \Pi_2) = 0 \quad \text{si} \quad H(\Pi_1) \cdot H(\Pi_2) = +\infty$$

rend équivalentes, pour tout couple $(\Pi_1, \Pi_2) \in E \times E$, les propositions

" Π_1 est H-indépendante de Π_2 " et " $H(\Pi_1; \Pi_2) = 0$ ".

Quand les expressions sont définies, nous avons aussi

$$H(\Pi_1; \Pi_2) = H(\Pi_1) - H(\Pi_1/\Pi_2)$$

$$H(\Pi_1; \Pi_2) = H(\Pi_2) - H(\Pi_2/\Pi_1)$$

De l'inégalité : $H(\Pi_1 \cap \Pi_2) \geq \sup(H(\Pi_1), H(\Pi_2))$

on déduit : $H(\Pi_1; \Pi_2) \leq \inf(H(\Pi_1), H(\Pi_2)).$

VI.1.3. - INFORMATION MOYENNE SUR E.

Comme nous l'avons déjà remarqué dans un chapitre précédent (V.2), il semble que, d'une manière générale, les axiomes 1 et 2 complétés par l'axiome 3 (qui est plutôt une définition) donnent des conditions nécessaires mais qui sont insuffisantes pour définir concrètement des informations. B. FORTE P. BENVENUTI et d'autres auteurs dans les articles précédemment cités, ont généralisé aux partitions les notions du type composabilité. Dans la suite, nous utiliserons plutôt des constructions du type information moyenne.

Etant donné une probabilité P et une information I au sens de J. KAMPE DE FERIET (I.1.1.) sur (Ω, S) , *l'information moyenne apportée par une partition de Ω est une information au sens que nous venons de définir (VI.1.2.).*

En effet, il est facile de montrer que la quantité

$$H(\Pi) = \sum_{A \in \Pi} P(A) \cdot I(A)$$

vérifie les axiomes 1 et 2 ; de plus $H(\{\Omega\}) = 0$. Avec cette définition nous avons la propriété supplémentaire suivante : *si tous les éléments A d'une partition Π_1 sont I -indépendants des éléments B d'une partition Π_2 , les partitions Π_1 et Π_2 sont H -indépendantes (l'inverse n'est pas nécessairement vrai).*

Quand les quantités qui interviennent sont définies, nous avons aussi

$$H(\Pi_1 / \Pi_2) = \sum_{(A,B) \in \Pi_1 \times \Pi_2} P(A \cap B) \cdot I(A/B)$$

$$H(\Pi_1; \Pi_2) = \sum_{(A,B) \in \Pi_1 \times \Pi_2} P(A \cap B) \cdot I(A; B) .$$

VI.1.4. - LA FORMULE DE SHANNON.

Nous pensons que le succès de la théorie classique de l'information tient aux propriétés particulières de l'information moyenne calculée par la formule de SHANNON. Essayons d'expliciter ces propriétés. Etant donné une mesure de probabilité P sur (Ω, S) , on sait que l'information apportée par une partition, au sens de SHANNON, s'écrit, à un coefficient multiplicatif (strictement positif) près :

$$H^*(\Pi) = \sum_{A \in \Pi} P(A) \cdot I^*(A)$$

ou $I^*(A) = \text{Log} \frac{1}{P(A)}$ est l'information apportée par un événement quelconque de (Ω, S) . $H^*(\Pi)$ est une quantité finie même si certains éléments A sont de probabilité nulle car $x \cdot \text{Log} \frac{1}{x}$ tend vers zéro avec x . Nous attachons le symbole $*$ aux informations I et H ainsi définies.

Il est facile de montrer [3], [27], [39] que $H^*(\Pi)$ est une information moyenne au sens du paragraphe VI.1.2. et qui possède en plus les propriétés suivantes :

$$H^*(\Pi_1 \cap \Pi_2) = H^*(\Pi_1) + H^*(\Pi_2) \quad <==>$$

$$\forall A \in \Pi_1, \forall B \in \Pi_2 \quad I^*(A \cap B) = I^*(A) + I^*(B)$$

c'est-à-dire que la H^* -indépendance en information des partitions Π_1 et Π_2 est équivalente à la I^* -indépendance en information des éléments de Π_1 et Π_2 (ce qui est aussi équivalent à l'indépendance en probabilité des éléments de Π_1 et Π_2).

On montre aussi que :

$$H^*(\Pi_1 \cap \Pi_2) \leq H^*(\Pi_1) + H^*(\Pi_2)$$

ce qui est équivalent à

$$H^*(\Pi_1; \Pi_2) = \sum_{(A,B) \in \Pi_1 \times \Pi_2} P(A \cap B) \cdot \log \frac{P(A \cap B)}{P(A) \cdot P(B)} \geq 0.$$

Pour tout $\Pi'_1 \leq \Pi_1$ et $\Pi'_2 \leq \Pi_2$ on a :

$$H^*(\Pi_1; \Pi_2) \leq H^*(\Pi'_1; \Pi'_2).$$

Nous avons donc pour $\Pi'_2 \leq \Pi_2$:

$$0 \leq H^*(\Pi_1; \Pi_2) \leq H^*(\Pi_1; \Pi'_2) \leq H^*(\Pi_1; \Pi_1) = H^*(\Pi_1)$$

avec dans cette dernière série d'inégalités

. $H^*(\Pi_1; \Pi_2) = 0$ si et seulement si Π_1 et Π_2 sont H^* -indépendantes.

. $H^*(\Pi_1; \Pi_2) = H^*(\Pi_1)$ si et seulement si $\Pi_2 \leq \Pi_1$

(en supposant que $(A,B) \in \Pi_1 \times \Pi_2 \implies P(A) \cdot P(B) > 0$)

Essayons d'interpréter ces résultats: pour tout Π_1 fixé, l'application $H^*(\Pi_1; \cdot)$ est une information sur $(E, \mathcal{L})$ au sens où nous l'avons définie dans le paragraphe VI.1.2., et la quantité $H^*(\Pi_1; \Pi_2)$ mesure ce que la connaissance de Π_1 nous apprend sur Π_2

$H^*(\Pi_1; \Pi_2) = 0$ si Π_1 est indépendante de Π_2

$H^*(\Pi_1; \Pi_2) = H^*(\Pi_1)$ si Π_2 est "plus fine" ou égale à Π_1 .

$H^*(\Pi_1; \Pi_2)$ est une quantité positive, inférieure à $H^*(\Pi_1)$, mais d'autant plus grande que la partition Π_2 se "rapproche" de la partition Π_1 .

Donnons un exemple : partitionnons l'ensemble Ω des livres d'une bibliothèque, d'une part par genre : $\Pi_1 = \{\text{Livres de mathématiques, de physique, d'histoire, ...}\}$; d'autre part par rayon : $\Pi_2 = \{\text{Livres du premier rayon, du deuxième rayon, ...}\}$; il peut être utile de savoir quantifier l'information moyenne que l'on apporte en indiquant le rayon sur lequel se trouve le livre qui nous intéresse, ou en donnant le genre de livre qui nous occupe. Ce n'est pas notre propos ; ce que nous quantifions c'est l'information $H^*(\Pi_1; \Pi_2)$ qu'une des deux partitions nous apporte sur l'autre partition :

- Si les livres sont répartis complètement au hasard dans la bibliothèque $H^*(\Pi_1; \Pi_2) = 0$ est nul ou voisin de zéro.

- Si les livres sont rigoureusement rangés : les livres de mathématiques sur tel rayon, les livres d'histoire sur tel rayon, etc..., $H^*(\Pi_1; \Pi_2)$ est maximum.

- Si les livres sont à peu près rangés, $H^*(\Pi_1; \Pi_2)$ doit avoir une valeur intermédiaire d'autant plus grande que les livres sont mieux rangés.

VI.1.5. - CONCLUSIONS.

Nous allons d'abord montrer que l'information mutuelle entre les deux partitions (nous dirons aussi l'information mutuelle entre les deux observateurs), définie à partir de l'information du chapitre IV, permet de retrouver toutes les propriétés de la formule de SHANNON. Ensuite, grâce à ces résultats et aux propriétés démontrées dans les chapitres II et III nous pourrions définir une information mutuelle entre deux groupes d'observateurs O_1 et O_2 qui généralise la quantité $H^*(\Pi_1; \Pi_2)$ et qui possède toutes les

propriétés intéressantes que nous venons de rappeler. L'interprétation de ces quantités nous permettra de résoudre en partie les problèmes que nous avons soulevés dans l'introduction du chapitre I.

Résumons-nous (en termes de partition) ; il est souhaitable que l'information H obtenue sur $(E, \mathcal{A})$ et qui est définie en liaison avec l'information I sur $(\Omega, \mathcal{A})$, possède les propriétés suivantes :

La H -indépendance en information des partitions Π_1 et Π_2 est équivalente à la I -indépendance en information des éléments $A \in \Pi_1$ et $B \in \Pi_2$.

Etant donné une partition Π_1 , l'information mutuelle $H(\Pi_1; \cdot)$ définit sur $(E, \mathcal{A})$ une information au sens du paragraphe VI.1.2. telle que

$$\forall \Pi_2 \in \mathcal{E} \quad 0 \leq H(\Pi_1; \Pi_2) \leq H(\Pi_1; \Pi_1) = H(\Pi_1)$$

avec

- $\cdot \quad H(\Pi_1; \Pi_2) = 0 \quad \text{si et seulement si la partition } \Pi_1 \text{ est } H\text{-indépendante de la partition } \Pi_2.$
- $\cdot \quad H(\Pi_1; \Pi_2) = H(\Pi_1) \quad \text{si et seulement si } \Pi_2 \leq \Pi_1.$

VI.2 - INFORMATION APPORTÉE PAR UN COUPLE D'OBSERVATEURS.

Les couples de partitions que nous voulons comparer ne sont pas quelconques ; nous allons associer à chaque observateur la partition qu'il distingue dans son espace propre, au sens ou nous l'avons défini au chapitre II, et les couples de partitions que nous utiliserons vérifieront donc la propriété suivante :

$$\forall A \in \Pi_1, \forall B \in \Pi_2 : A \cap B \neq \emptyset$$

c'est-à-dire que ces partitions seront a priori indépendantes en structure (I.1.1.).

Imaginons un exemple simple qui va nous permettre de concrétiser nos idées. Un biologiste dispose de deux tests a et b pour obtenir un certain renseignement sur une population de bactéries. Nous supposons que le test b qui est coûteux, difficile à mettre en oeuvre, donne sans erreur le résultat recherché et que le test a , simple et peu onéreux, donne une réponse moins fiable à la question du biologiste. Après avoir expérimenté les deux tests sur une population témoin de 100 bactéries, le biologiste qui n'utilisera plus que le test a , nous demande de définir à partir du tableau ci-dessous quelles informations le test a lui apporte sur le vrai résultat (inconnu), c'est-à-dire sur le résultat du test b si on pouvait l'effectuer.

Notons $\Omega_a = \{a_+, a_-\}$ et $\Omega_b = \{b_+, b_-\}$ les résultats possibles de chacun des tests (+ = réagit, - = ne réagit pas), nous avons par exemple :

$\Omega_a \backslash \Omega_b$	a_+	a_-	TOTAL
b_+	50	10	60
b_-	5	35	40
TOTAL	55	45	100

Tableau 1 :
Fréquence en pourcentage

Commençons par faire une synthèse des résultats dont nous avons besoin en explicitant en détail les rapports existant entre les notations des chapitres précédents et cette étude.

VI.2.1. - APPLICATIONS DES RESULTATS DES PARAGRAPHEs II.1 et IV.1.

Soit un ensemble $\mathcal{O}$ composé de deux observateurs a et b et $F = P(\mathcal{O})$. Associons respectivement à chacun des observateurs a et b les espaces probabilisables (Ω_a, S_a) et (Ω_b, S_b) ; les tribus S_a et S_b de parties de Ω_a et Ω_b ne comportent qu'un nombre fini d'éléments. Définissons l'espace mesurable $(\hat{\Omega}, \hat{S})$ produit de ces deux espaces :

$$\hat{\Omega} = \Omega_a \times \Omega_b$$

$$\hat{S} = S_a \otimes S_b$$

Afin de simplifier l'écriture et avec les conventions du chapitre IV, Nous notons A (resp. B) la sous-tribu de $\hat{S}$ des cylindres de base dans S_a (resp. S_b) (c'est-à-dire $A = \hat{S}_a$ (resp. $B = \hat{S}_b$), avec les notations de I.1.3.) et A (resp. B) les éléments de A (resp. B). Pour tout $C \in \hat{S}$ nous notons C_a et C_b les cylindres de A et B dont la base est la projection de C respectivement sur (Ω_a, S_a) et (Ω_b, S_b) .

Soit I une information donnée sur $(\hat{\Omega}, \hat{S})$ au sens I.1.1., nous avons montré (II.1) que l'application J définie ci-dessous à partir de I est une information sur $(\hat{\Omega} \times \mathcal{O}, \hat{S} \times F)$ au sens I.2.2., I.2.6.

$$\begin{aligned} \forall C \in \hat{S} \quad J(C, \emptyset) &= 0 \\ J(C, \{a\}) &= I(C_a) \\ J(C, \{b\}) &= I(C_b) \\ J(C, \mathcal{O}) &= I(C) \end{aligned}$$

En particulier nous avons

$$\forall (A, B) \in A \times B$$

$$J((A \cap B), \emptyset) = 0$$

$$J((A \cap B), \{a\}) = I(A)$$

$$J((A \cap B), \{b\}) = I(B)$$

$$J((A \cap B), \emptyset) = I(A \cap B)$$

VI.2.2. - APPLICATION DES RESULTATS DU PARAGRAPHE II.2.

L'espace $(\hat{\Omega}, \hat{S})$ que nous venons de définir est ce que nous avons appelé l'espace des observations. Notons Ω l'ensemble des 100 bactéries de l'exemple étudié, $S = P(\Omega)$. A chaque élément $\omega \in \Omega$ associons les points $T_a(\omega) \in \Omega_a$ et $T_b(\omega) \in \Omega_b$ indiquant la réaction de la bactérie ω aux tests a et b . L'application mesurable T définie ainsi sur (Ω, S) à valeurs dans $(\hat{\Omega}, \hat{S})$ formalise donc la réalisation de l'expérience.

Nous avons sur (Ω, S) une mesure de probabilité P qui est naturelle ; c'est celle qui affecte à chaque individu $\omega \in \Omega$ le même poids. La variable aléatoire T définie sur (Ω, S, P) à valeurs dans $(\hat{\Omega}, \hat{S})$ permet donc de définir une probabilité P_T sur $(\hat{\Omega}, \hat{S})$. En fait dans la pratique on ne distingue pas en général la probabilité P sur (Ω, S) et P_T sur $(\hat{\Omega}, \hat{S})$ et l'on assimile par exemple la probabilité de l'ensemble des bactéries de $\omega \in \Omega$ qui ont répondu *positivement* au test a , à la probabilité de la réponse positive du test a . Le tableau 1 nous donne la probabilité sur $(\hat{\Omega}, \hat{S})$ que nous étudions. Ce que nous appelons probabilité est en fait une pondération positive de masse totale égale à 1 sur (Ω, S) .

VI.2.3. - APPLICATION DES RESULTATS DU CHAPITRE IV.

Nous avons supposé dans le chapitre IV (en reprenant les notations de VI.2.1.) qu'il existe une mesure de probabilité P sur $(\hat{\Omega}, \hat{\mathcal{S}})$ telle que pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$ et $A \neq \emptyset, B \neq \emptyset$ nous avons

$$P(A) > 0, \quad P(B) > 0$$

et seulement $P(A \cap B) \geq 0$.

Nous avons émis les hypothèses suivantes : d'une part (H,2) pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$:

$$I'(A \cap B) = F(P(A), P(B), P(A \cap B)),$$

l'information I' peut être calculée à partir de la probabilité P ; d'autre part (H.1) pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$ et pour toute partition finie Π_a, Π_b respectivement $\mathcal{A}$ -mesurable et $\mathcal{B}$ -mesurable de $\hat{\Omega}$, nous avons :

$$I'(A) = \sum_{B' \in \Pi_b} P(B') \cdot I'(A/B')$$

$$I'(B) = \sum_{A' \in \Pi_a} P(A') \cdot I'(B/A')$$

Nous montrons dans le théorème IV.3.7. qu'une condition nécessaire et suffisante pour que les hypothèses H_1 et H_2 soient vérifiées, est que pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$, nous ayons :

$$I'(A \cap B) = I'(A) + I'(B) - K \cdot \left(\frac{P(A \cap B)}{P(A) \cdot P(B)} - 1 \right)$$

avec

$$I'(A) = f_1(P(A))$$

$$I'(B) = f_2(P(B))$$

K est un nombre positif ou nul et les applications f_1 et f_2 sont soumises aux conditions énoncées dans le théorème IV.3.7.

En particulier pour tout $A \in \mathcal{A}$

$$I'(A) \geq K \cdot I^0(A) = K \cdot \left(\frac{1}{P(A)} - 1 \right)$$

De même pour $B \in \mathcal{B}$:

$$I'(B) \geq K \cdot I^0(B) = K \cdot \left(\frac{1}{P(B)} - 1 \right) .$$

Il vient alors :

$$I'(A \cap B) \geq K \cdot I^0(A \cap B)$$

où I^0 est l'information au sens I.1.1. sur $(\hat{\Omega}, \hat{\mathcal{S}})$ définie par

$$I^0(A \cap B) = I^0(A) + I^0(B) - \left(\frac{P(A \cap B)}{P(A) \cdot P(B)} - 1 \right) = \frac{P(A) + P(B) - P(A \cap B) - P(A) \cdot P(B)}{P(A) \cdot P(B)}$$

Notons que, pour les informations mutuelles, nous avons :

$$I'(A; B) = K \cdot I^0(A; B) = K \cdot \left(\frac{P(A \cap B)}{P(A) \cdot P(B)} - 1 \right)$$

La constante K est quelconque, positive ou nulle. *Nous allons dans la suite supposer que la constante K est strictement positive et égale à 1.* En effet, nous avons montré dans le paragraphe IV.4 qu'il existe des relations intéressantes entre les notions d'indépendance et de dépendance pour la probabilité P et pour l'information I' quand la constante K est strictement positive, en particulier :

$$I'(A \cap B) = I'(A) + I'(B) \iff P(A \cap B) = P(A).P(B) ,$$

$$I^0(A \cap B) = I^0(A) + I^0(B) \iff P(A \cap B) = P(A).P(B)$$

Par ailleurs, il est clair que si l'on multiplie ou divise une information I au sens de J. KAMPE DE FERIET par une constante K strictement positive, on obtient à nouveau une information de même type. *Nous attachons le symbole I' à l'information solution générale du chapitre IV et I^0 à la solution particulière que nous venons de rappeler.*

VI.2.4. - INFORMATIONS MUTUELLES ENTRE DEUX OBSERVATEURS.

Soit Π_a une partition finie, A -mesurable, de $\hat{\Omega}$; calculons l'information moyenne apportée par cette partition avec les notations du paragraphe précédent, il vient :

$$H'(\Pi_a) = \sum_{A \in \Pi_a} P(A).I'(A)$$

$$H^0(\Pi_a) = \sum_{A \in \Pi_a} P(A).I^0(A) = \text{Card}(\Pi_a) - 1 ,$$

de plus, il est clair que : $H'(\Pi_a) \geq H^0(\Pi_a)$. Nous avons des résultats équivalents pour toute partition finie Π_b , B -mesurable, de $\hat{\Omega}$.

Soient Π_a et Π_b deux partitions finies respectivement A et B -mesurables de $\hat{\Omega}$; nous avons :

$$H'(\Pi_a \cap \Pi_b) = \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A \cap B).I'(A \cap B)$$

$$H^0(\Pi_a \cap \Pi_b) = \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A \cap B) \cdot I^0(A \cap B).$$

Il est clair que H' et H^0 sont des informations moyennes, au sens indiqué dans les paragraphes VI.1.2. et VI.1.3. sur l'ensemble des partitions de $\hat{\Omega}$ de la forme $\Pi_a \cap \Pi_b$. Nous pouvons écrire aussi :

$$H'(\Pi_a \cap \Pi_b) = H'(\Pi_a) + H'(\Pi_b) - H'(\Pi_a; \Pi_b)$$

$$H^0(\Pi_a \cap \Pi_b) = H^0(\Pi_a) + H^0(\Pi_b) - H^0(\Pi_a; \Pi_b)$$

avec

$$H'(\Pi_a; \Pi_b) = H^0(\Pi_a; \Pi_b) = \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A \cap B) \cdot \left(\frac{P(A \cap B)}{P(A) \cdot P(B)} - 1 \right).$$

Cette information mutuelle entre les partitions Π_a et Π_b est une quantité bien connue ; elle s'écrit aussi :

$$\begin{aligned} H_0(\Pi_a; \Pi_b) &= \sum_{(A,B) \in \Pi_a \times \Pi_b} \frac{1}{P(A) \cdot P(B)} (P(A \cap B) - P(A) \cdot P(B))^2 \\ &= \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A) \cdot P(B) \left(\frac{P(A \cap B)}{P(A) \cdot P(B)} - 1 \right)^2 \\ &= \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A) \cdot P(B) \left(\left(\frac{P(A \cap B)}{P(A) \cdot P(B)} \right)^2 - 1 \right) \end{aligned}$$

C'est la distance dite du CHI - 2 entre la loi de probabilité P sur $(\hat{\Omega}, \hat{S})$ et la loi de probabilité sur $(\hat{\Omega}, \hat{S})$ définie par le produit des lois marginales de P sur $(\hat{\Omega}, \hat{A})$ et $(\hat{\Omega}, \hat{B})$; en supposant que A et B sont les tribus engendrées par Π_a et Π_b .

On sait [3] que les propriétés de concavité de la fonction $x^2 - 1 = \left(\frac{P(A \cap B)}{P(A).P(B)} \right)^2 - 1$ permettent de montrer facilement que $H^0(\Pi_a; \Pi_b)$

possède les propriétés suivantes :

• Pour tout couple de partitions finies Π_a et Π_b , respectivement A et B-mesurables de $\hat{\Omega}$ on a :

$$H^0(\Pi_a; \Pi_b) \geq 0$$

avec égalité si et seulement si

$$P(A \cap B) = P(A).P(B) \quad \text{pour tout } (A, B) \in \Pi_a \times \Pi_b$$

c'est-à-dire que la H^0 -indépendance en information des partitions Π_a et Π_b est équivalente à la I^0 -indépendance en information des éléments $A \in \Pi_a$ et $B \in \Pi_b$.

• Soient Π_a, Π'_a (respectivement Π_b, Π'_b) deux partitions finies A-mesurables (resp. B-mesurables) de $\hat{\Omega}$

$$\left. \begin{array}{l} \Pi'_a \leq \Pi_a \\ \Pi'_b \leq \Pi_b \end{array} \right\} \implies H^0(\Pi'_a; \Pi'_b) \geq H^0(\Pi_a; \Pi_b)$$

en particulier, étant donné Π_a , l'information mutuelle $H^0(\Pi_a; \cdot)$ définit sur les partitions B-mesurables de $\hat{\Omega}$ une information au sens du paragraphe VI.1.2. :

$$\Pi'_b \leq \Pi_b \implies 0 \leq H^0(\Pi_a; \Pi_b) \leq H^0(\Pi_a; \Pi'_b) .$$

On a enfin :

$$H^0(\Pi_a; \Pi_b) \leq H^0(\Pi_a)$$

avec égalité si et seulement si

$$\forall B \in \Pi_b, \exists A \in \Pi_a : P(A \cap B) = P(B)$$

c'est-à-dire en fait si $\Pi_b \leq \Pi_a$, non pas dans $(\hat{\Omega}, \hat{S})$ où cette relation n'a pas de sens, mais dans l'espace (Ω, S) sur lequel on a défini la mesure de probabilité : $T^{-1}(\Pi_b) \leq T^{-1}(\Pi_a)$ à des ensembles P-négligeables près.

VI.2.5. - CONCLUSIONS.

Nous voyons donc que les informations I^0 et H^0 sur $(\hat{\Omega}, \hat{S})$ (ainsi que I' et H' , car $H'(\Pi_a; \Pi_b) = H^0(\Pi_a; \Pi_b)$) possèdent toutes les propriétés que nous avons énoncées dans le paragraphe VI.1.5.

On trouvera dans le chapitre intitulé "Théorie de l'information et classification d'après un tableau de contingence" du livre de J.P. BENZECRI et collaborateurs [3] une étude détaillée et une comparaison approfondie des quantités $H^0(\Pi_1; \Pi_2)$ et $H^*(\Pi_1; \Pi_2)$ ($H^*(\Pi_1; \Pi_2)$: information mutuelle au sens de SHANNON). En conclusion de ce paragraphe et en nous inspirant de [3] nous pouvons faire les remarques suivantes :

* Les informations I^0, H^0 que nous venons de définir et les informations I^*, H^* au sens de SHANNON possèdent à notre sens rigoureusement les mêmes propriétés, au moins quant à la notion d'information mutuelle. A priori toutes les applications de la théorie classique qui utilisent la notion d'information mutuelle peuvent, semble-t-il, être reprises dans le cadre que nous venons de définir.

* On sait que l'on peut obtenir la forme de l'information au sens de SHANNON en partant des axiomes suivants :

$$I^*(A) = f(P(A))$$

$$I^*(A \cap B) = I^*(A) + I^*(B) \iff P(A \cap B) = P(A).P(B).$$

l'équation fonctionnelle ainsi introduite implique $f(.) = -\text{Log}(.)$.

En fait dans le chapitre IV nous supposons aussi (H.2) que l'information I^0 est une fonction de la probabilité mais de la manière suivante :

$$I^0(A \cap B) = F(P(A), P(B), P(A \cap B))$$

(A et B appartenant à des espaces indépendants en structure

$$(A, B) \in A \times B \implies A \cap B \neq \emptyset$$

et nous obtenons comme propriété que :

$$I^0(A \cap B) = I^0(A) + I^0(B) \iff P(A \cap B) = P(A) \cdot P(B).$$

Pour déterminer une forme possible pour I^0 nous introduisons dans le chapitre IV une autre hypothèse (H.1.) ce qui nous conduit comme dans le cas de SHANNON à résoudre une équation fonctionnelle qui nous donne les solutions que nous venons d'étudier.

* Remarquons enfin qu'il est facile de trouver d'autres formes d'informations I et H sur $(\hat{\Omega}, \hat{S})$ qui vérifient l'ensemble de ces propriétés.

VI.2.6. - EXEMPLE.

Reprenons l'exemple de l'introduction de ce paragraphe et notons $A = \{a_+\}$, $A^C = \{a_-\}$, $B = \{b_+\}$, $B^C = \{b_-\}$; nous obtenons par application des formules précédentes aux données du tableau de fréquence suivant :

$\Omega_a \backslash \Omega_b$	B	B^C	TOTAL
A	50	5	55
A^C	10	35	45
TOTAL	60	40	100

50 + 5 = 55
10 + 35 = 45
55 + 45 = 100

a) I^0 et H^0 .

$$\cdot I^0(A) = 0,81 ; I^0(A^C) = 1,22 ; I^0(B) = 0,67 ; I^0(B^C) = 1,5.$$

$$\cdot I^0(A \cap B) = 0,97 ; I^0(A \cap B^C) = 3,09 ; I^0(A^C \cap B) = 2,52 ;$$

$$I^0(A^C \cap B^C) = 1,78$$

$$\cdot I^0(A; B) = 0,51 ; I^0(A; B^C) = -0,78 ; I^0(A^C; B) = -0,63 ;$$

$$I^0(A^C; B^C) = 0,94$$

$$\cdot H^0(\pi_a) = H^0(\pi_b) = 1$$

$$\cdot H^0(\pi_a \cap \pi_b) = 1,51$$

$$\cdot H^0(\pi_a; \pi_b) = 0,49.$$

Il est facile de vérifier que si $P(A) = P(B) = P(A \cap B) = 0,60$ et $P(A^C) = P(B^C) = 0,40$, c'est-à-dire si $\Pi_a = \Pi_b$ sur la population Ω de bactéries étudiées, on a alors $H^0(\Pi_a; \Pi_b) = H^0(\Pi_a) = 1$.

De même si les partitions Π_a et Π_b sont indépendantes en probabilité, on obtient $H^0(\Pi_a; \Pi_b) = 0$.

b) I^* et H^* .

$$\cdot I^*(A) = 0,86 ; I^*(A^C) = 1,15 ; I^*(B) = 0,74 ; I^*(B^C) = 1,32$$

$$\cdot I^*(A \cap B) = 1 ; I^*(A \cap B^C) = 4,32 ; I^*(A^C \cap B) = 3,32 ; I^*(A^C \cap B^C) = 1,51$$

$$\cdot I^*(A; B) = 0,60 ; I^*(A; B^C) = -2,14 ; I^*(A^C; B) = -1,43 ; I^*(A^C; B^C) = 0,96$$

$$\cdot H^*(\Pi_a) = 0,99 ; H^*(\Pi_b) = 0,97 ;$$

$$\cdot H^*(\Pi_a \cap \Pi_b) = 1,58$$

$$\cdot H^*(\Pi_a; \Pi_b) = 0,39$$

Si $P(A) = P(B) = P(A \cap B) = 0,60$ on a alors

$$H^*(\Pi_a; \Pi_b) = H^*(\Pi_a) = 0,97.$$

De même si les partitions Π_a et Π_b sont indépendantes en probabilité, on obtient

$$H^*(\Pi_a; \Pi_b) = 0$$

c) Remarque.

Les quantités $H^0(\Pi_a; \Pi_b)$ et $H^*(\Pi_a; \Pi_b)$ répondent à la question posée dans l'introduction du paragraphe VI.2. Quel sens donner aux valeurs obtenues ?

Si n est l'effectif total de la population étudiée

$$n \cdot H^0(\Pi_a; \Pi_b) = \sum_{(A,B) \in \Pi_a \times \Pi_b} \frac{(n \cdot P(A \cap B) - n \cdot P(A) \cdot P(B))^2}{n \cdot P(A) \cdot P(B)}$$

$$n.H^0(\Pi_a; \Pi_b) = \sum_{i,j} \frac{(O_{ij} - C_{ij})^2}{C_{ij}}$$

c'est-à-dire la valeur calculée habituellement dans le test du CHI - 2 (O_{ij} : effectifs observés ; C_{ij} : effectifs calculés) qui sert à étudier l'indépendance en probabilité des partitions Π_a et Π_b . On voit que, pour un tableau de fréquences donné, si l'on prend un effectif n suffisamment grand, on pourra toujours démontrer par ce test la non-indépendance en probabilité des partitions Π_a et Π_b dès que $H^0(\Pi_a; \Pi_b) > 0$. Notre point de vue n'est pas celui-là ; il ne s'agit pas pour nous de démontrer l'indépendance ou la non-indépendance de Π_a et Π_b mais de mesurer leur degré de dépendance. S'il fallait faire une étude statistique de $H^0(\Pi_a; \Pi_b)$ (ou de n fois $H^0(\Pi_a; \Pi_b)$) notre hypothèse nulle ne serait pas l'indépendance de Π_a et Π_b c'est-à-dire $H^0(\Pi_a; \Pi_b) = 0$, mais au contraire $H^0(\Pi_a; \Pi_b) = H^0(\Pi_a)$, c'est-à-dire dans l'exemple précédent, d'étudier si la valeur $H^0(\Pi_a; \Pi_b) = 0,49$ diffère significativement, non pas de zéro, mais de $H^0(\Pi_a) = 1$.

L'étude de ce type de problème (par les méthodes des statistiques classiques ou par des méthodes de simulation) apparaît indispensable pour certaines applications ; nous verrons dans le paragraphe VII.1.1., un exemple de ce type traité par une méthode de simulation.

VI.3 - INFORMATION K APPOREE SUR UNE PARTITION PAR UN ENSEMBLE FINI D'OBSERVATEURS.

Reprenons les notations du paragraphe VI.2.4. : soient Π_a et Π_b deux partitions finies, respectivement A et B mesurables de $\hat{\Omega}$; nous avons défini :

$$H(\Pi_a) = \sum_{A \in \Pi_a} P(A).I(A)$$

$$H(\Pi_a \cap \Pi_b) = \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A \cap B).I(A \cap B).$$

Notons $\Pi = \Pi_a \cap \Pi_b$ et calculons $K(\Pi, \{a\})$ l'information moyenne apportée sur Π par l'observateur $a \in \mathcal{O}$ c'est-à-dire :

$$K(\Pi, \{a\}) = \sum_{C \in \Pi} P(C) \cdot J(C, \{a\}).$$

Nous savons que $C \in \Pi$, $C \neq \emptyset$ est de la forme $A \cap B$, $A \in \Pi_a$ et $B \in \Pi_b$ et (VI.2.1.) que :

$$J((A \cap B), \{a\}) = I(A)$$

ce qui entraîne

$$K(\Pi, \{a\}) = \sum_{(A,B) \in \Pi_a \times \Pi_b} P(A \cap B) \cdot I(A) = \sum_{A \in \Pi_a} P(A) \cdot I(A)$$

$$K(\Pi, \{a\}) = H(\Pi_a).$$

De même

$$K(\Pi, \{b\}) = H(\Pi_b)$$

$$K(\Pi, \{a,b\}) = H(\Pi)$$

et évidemment $K(\Pi, \emptyset) = 0$.

Sur la même partition Π de $\hat{\Omega}$, l'information apportée par un ensemble vide d'observateurs est nulle ; l'information apportée par l'observateur a est égale à l'information moyenne de la partition perçue par cet observateur : Π_a , l'information apportée par les deux observateurs est égale à l'information moyenne apportée par la partition Π elle-même. Il est clair que l'application K est une information moyenne qui généralise l'information moyenne H sur les

partitions de la même manière que l'information J généralise l'information I sur les événements. Conformément aux considérations générales de I.1.3. en nous inspirant des paragraphes I.2. et VI.1.2., nous sommes amenés à proposer les définitions du paragraphe suivant.

VI.3.1. - INFORMATION K SUR $E \times F$.

Soit $\emptyset$ un ensemble fini d'observateurs, $F = P(\emptyset)$. L'information apportée sur une partition finie $\Pi \in E$ de l'ensemble Ω , par un groupe $F \in F$ d'observateurs, notée $K(\Pi, F)$, est définie par une application vérifiant les axiomes suivants

$$A.1. \quad K : E \times F \longrightarrow \mathbb{R}^+$$

A.2. L'information K est monotone sur $E \times F$

$$\left. \begin{array}{l} (\Pi_1, \Pi_2) \in E \times E \text{ et } \Pi_1 \leq \Pi_2 \\ (F_1, F_2) \in F \times F \text{ et } F_1 \supseteq F_2 \end{array} \right\} \implies$$

$$K(\Pi_1, F_1) \geq K(\Pi_2, F_2).$$

A.3. Deux éléments $(\Pi_i, F_i) \in E \times F$, $i = 1, 2$ sont indépendants pour l'information K si

$$K((\Pi_1 \cap \Pi_2), (F_1 \cup F_2)) = K(\Pi_1, F_1) + K(\Pi_2, F_2).$$

Nous admettrons comme "valeurs universelles"

$$\forall \Pi \in E \quad K(\Pi, \emptyset) = 0$$

$$\forall F \in F \quad K(\{\Omega\}, F) = 0$$

L'information conditionnelle fournie par l'élément $(\Pi_1, F_1) \in E \times F$ connaissant l'information apportée par l'élément $(\Pi_2, F_2) \in E \times F$ est définie par :

$$K((\Pi_1, F_1)/(\Pi_2, F_2)) = K((\Pi_1 \cap \Pi_2), (F_1 \cup F_2)) - K(\Pi_2, F_2)$$

si $K(\Pi_2, F_2) < +\infty$.

L'information mutuelle apportée par les éléments $(\Pi_i, F_i) \in E \times F$, $i = 1, 2$ est définie par :

$$K((\Pi_1, F_1); (\Pi_2, F_2)) = K(\Pi_1, F_1) + K(\Pi_2, F_2) - K((\Pi_1 \cap \Pi_2), (F_1 \cup F_2))$$

si $K(\Pi_1, F_1) < +\infty$ et $K(\Pi_2, F_2) < +\infty$.

Quand les expressions sont définies nous avons

$$K((\Pi_1, F_1); (\Pi_2, F_2)) = K(\Pi_1, F_1) - K((\Pi_1, F_1)/(\Pi_2, F_2))$$

$$K((\Pi_1, F_1); (\Pi_2, F_2)) = K(\Pi_2, F_2) - K((\Pi_2, F_2)/(\Pi_1, F_1)).$$

Si $\Pi_1 = \Pi_2 = \Pi$ nous écrirons aussi $K(\Pi, (F_1, F_2))$ pour $K((\Pi, F_1); (\Pi, F_2))$, notons enfin que $K(\Pi, (F_1, F_1)) = K(\Pi, F_1)$.

VI.3.2. - INFORMATION MOYENNE K SUR $E \times F$.

Etant donné une probabilité P sur (Ω, S) et une information J (I.2.2. - I.2.6.) sur $(\Omega \times \emptyset, S \times F)$, $F_\emptyset = \{\emptyset\}$, l'information moyenne apportée par un ensemble $F \in \mathcal{F}$ d'observateurs sur une partition de Ω est une information K au sens où nous venons de la définir.

En effet, la quantité définie par

$$\forall (\Pi, F) \in E \times F \quad K(\Pi, F) = \sum_{A \in \Pi} P(A) \cdot J(A, F)$$

vérifie les axiomes 1 et 2 ; de plus $\forall (\Pi, F) \in E \times F$

$$K(\Pi, \emptyset) = 0 \quad \text{et} \quad K(\{\Omega\}, F) = 0.$$

Nous avons les propriétés suivantes : étant donné $(\Pi_i, F_i) \in E \times F$, $i = 1, 2$, si tous les éléments $\{(A, F_1) | A \in \Pi_1\}$ sont J -indépendants des éléments $\{(B, F_2) | B \in \Pi_2\}$, les éléments $(\Pi_i, F_i) \in E \times F$, $i = 1, 2$, sont K -indépendants, (L'inverse n'est pas nécessairement vrai). Quand les quantités suivantes sont définies, nous avons aussi :

$$K((\Pi_1, F_1) / (\Pi_2, F_2)) = \sum_{(A, B) \in \Pi_1 \times \Pi_2} P(A \cap B) \cdot J((A, F_1) / (B, F_2))$$

$$K((\Pi_1, F_1); (\Pi_2, F_2)) = \sum_{(A, B) \in \Pi_1 \times \Pi_2} P(A \cap B) \cdot J((A, F_1); (B, F_2))$$

VI.3.3. - INFORMATION MUTUELLE ENTRE DEUX GROUPES D'OBSERVATEURS.

Nous avons vu au début de ce paragraphe VI.3 que pour $\emptyset = \{a, b\}$,

$\Pi = \Pi_a \cap \Pi_b$ nous avons :

$$K(\Pi, \{a\}) = H(\Pi_a) \quad ; \quad K(\Pi, \{b\}) = H(\Pi_b)$$

$$K(\Pi, \{a, b\}) = H(\Pi_a \cap \Pi_b)$$

nous avons donc

$$K((\Pi, \{a\}); (\Pi, \{b\})) = K(\Pi, (\{a\}; \{b\})) = H(\Pi_a; \Pi_b)$$

et nous pouvons interpréter alors les résultats énoncés à la fin des paragraphes VI.1.5. et VI.2.4. de la manière suivante : étant donné Π et l'observateur $a \in \emptyset$ la quantité

$$K(\Pi, (\{a\}; \{a\})) = H(\Pi_a)$$

définit l'information apportée (sur Π) par l'observateur a , et la quantité $K(\Pi, (\{a\}; \{b\}))$ qui est telle que :

$$0 \leq K(\Pi, (\{a\}; \{b\})) \leq K(\Pi, (\{a\}; \{a\})) = H(\Pi_a)$$

défini l'information apportée par l'observateur b sur l'information apportée par l'observateur a (pour l'observation de la partition Π) ; cette quantité est nulle si les observateurs a et b sont indépendants, elle est maximum si l'observateur b est "meilleur" que l'observateur a (on identifie évidemment, dans cet énoncé, l'observateur à la partition qu'il distingue).

D'une manière plus générale, pour un ensemble fini Ω quelconque d'observateurs, dans les paragraphes suivants, nous allons définir une information K telle que pour toute partition Π donnée et pour tous groupes F_1, F_2 d'observateurs de F convenablement choisis :

$$0 \leq K(\Pi, (F_1; F_2)) \leq K(\Pi, (F_1; F_1)) = K(\Pi, F_1)$$

afin que l'information mutuelle $K(\Pi, (F_1; F_2))$ soit une quantité nulle si les groupes d'observateurs F_1 et F_2 sont K -indépendants ; qu'elle soit égale à $K(\Pi, F_1)$ si les observateurs de F_2 apportent sur Π au moins l'information que les observateurs de F_1 apportent sur Π ; et que les valeurs de $K(\Pi, (F_1; F_2))$ soient d'autant plus grandes que les observateurs de F_2 résument mieux l'information des observateurs de F_1 pour la partition Π .

VI.4. - INFORMATIONS J et K SUR UN ESPACE PRODUIT.

VI.4.1. - RAPPEL DES NOTATIONS DU CHAPITRE II.

Examinons l'exemple de la page ci-après.

Le tableau $\Omega \times \Omega$ formalise la réalisation d'une enquête, d'une expérience... par exemple, un biologiste a réalisé sur une population Ω de 6 bactéries une série Ω de 4 tests. (Dans la réalité, on aurait évidemment quelques centaines ou quelques milliers d'individus $\omega \in \Omega$, étudiés suivant plusieurs

dizaines de tests $\xi \in \mathcal{O}$). La dernière ligne du tableau signifie par exemple que la bactérie ω_6 a réagi de la manière codée 2 au test α , 1 au test β et zéro aux tests γ et δ . Nous n'attachons aucun sens particulier aux symboles utilisés pour le codage ; pour le test α , nous avons codé les trois résultats possibles par 0,1,2 ; nous aurions pu utiliser les symboles α' , α'' , α''' : Il s'agit donc d'une simple convention. Dans cet exemple, l'ensemble $\mathcal{O}$ des tests (ou des caractères) joue le rôle que nous attribuons à l'ensemble des observateurs.

		$\mathcal{O}$			
		α	β	γ	δ
Ω	ω_1	0	0	0	0
	ω_2	0	1	0	1
	ω_3	0	1	1	1
	ω_4	1	1	1	0
	ω_5	2	0	1	1
	ω_6	2	1	0	0

D'une manière plus générale nous associons à chaque observateur ξ d'un ensemble fini $\mathcal{O}$ l'ensemble fini Ω_ξ des modalités du caractère ξ . Posons $S_\xi = P(\Omega_\xi)$ et notons (I.1.3.) :

$$\hat{\Omega} = \Omega_{\xi_1} \times \Omega_{\xi_2} \times \dots = \prod_{\mathcal{O}} \Omega_\xi$$

$$\hat{S} = S_{\xi_1} \otimes S_{\xi_2} \otimes \dots = \bigotimes_{\mathcal{O}} S_\xi$$

Dans ce paragraphe nous nous intéresserons essentiellement aux pavés mesurables de $(\hat{\Omega}, \hat{S})$ c'est-à-dire aux ensembles $\hat{C}$ de la forme :

$$\hat{C} = \prod_{\xi \in \mathcal{O}} C_{\xi} \quad \text{avec} \quad (C_{\xi})_{\xi \in \mathcal{O}} \in \prod_{\mathcal{O}} S_{\xi}$$

Dans l'exemple, nous avons :

$$\Omega_{\alpha} = \{0, 1, 2\}, \quad \Omega_{\beta} = \Omega_{\gamma} = \Omega_{\delta} = \{0, 1\}$$

$$\hat{\Omega} = \{0, 1, 2\} \times \{0, 1\}^3$$

$$\hat{C} = \{0, 1\} \times \Omega_{\beta} \times \{0\} \times \{1\} \quad \text{est un pavé de } \hat{S}.$$

La probabilité P sur $(\hat{\Omega}, \hat{S})$.

Le tableau $\Omega \times \mathcal{O}$ définit l'application T (II.2) qui associe à chaque bactérie $\omega \in \Omega$ le point $T(\omega) \in \hat{\Omega}$ représentatif des réponses de cette bactérie aux tests

Il y a sur $(\Omega, S), S = \mathcal{P}(\Omega)$, une mesure de probabilité P naturelle : celle qui associe à chaque point $\omega \in \Omega$ la même probabilité et l'application mesurable T définie sur (Ω, S) à valeurs dans $(\hat{\Omega}, \hat{S})$ permet de définir sur ce dernier espace une mesure de probabilité P_T de la manière classique :

$$\hat{C} \in \hat{S} \quad P_T(\hat{C}) = P(T^{-1}(\hat{C})) .$$

Dans l'exemple P_T est engendrée par

$$\forall x \in T(\Omega) \quad P_T(x) = \frac{1}{6}$$

$$\forall x \notin T(\Omega) \quad P_T(x) = 0$$

Dans la suite, nous travaillons toujours sur $(\hat{\Omega}, \hat{S})$ et nous ne distinguons pas P et P_T c'est-à-dire que nous identifierons dans la pratique le point $2 \in \Omega_\alpha$ avec l'ensemble des bactéries $\{\omega_5, \omega_6\}$ de Ω qui ont réagi de la manière codée 2 au test α . Nous supposons de plus que pour tout $\xi \in \mathcal{O}$ et tout $x_\xi \in \Omega_\xi$, $P(\{x_\xi\}) > 0$.

Il convient cependant d'être prudent pour cette identification. A tout sous-ensemble $\hat{C}$ de $\hat{S}$ il est facile de faire correspondre le sous-ensemble $T^{-1}(\hat{C})$ de Ω qui lui est associé

$$\begin{aligned}\hat{C} &= \{1, 2\} \times \{1\} \times \Omega_Y \times \{0\} \\ T^{-1}(\hat{C}) &= \{\omega_4, \omega_6\}.\end{aligned}$$

Mais ce même ensemble de bactéries peut être aussi défini à partir d'un autre pavé

$$\begin{aligned}\hat{C}' &= \{1, 2\} \times \{1\} \times \Omega_Y \times \Omega_\delta \\ T^{-1}(\hat{C}') &= \{\omega_4, \omega_6\}.\end{aligned}$$

Nous avons évidemment $P(\{\omega_4, \omega_6\}) = P_T(\hat{C}) = P_T(\hat{C}')$ mais comme $\hat{C} \subset \hat{C}'$ le cas échéant, l'information apportée par $\hat{C}$ pourrait être strictement plus grande que l'information apportée par $\hat{C}'$. Aussi, par convention, afin de faciliter les notations et en accord avec les résultats des paragraphes II.2 et II.4, quand nous parlerons dans la suite de l'information apportée par un ensemble $A \in \mathcal{S}$, cela signifiera en fait l'information apportée par le plus petit pavé de $\hat{S}$, noté $\hat{C}(A)$, qui est tel que $T^{-1}(\hat{C}(A))$ contient A , c'est-à-dire l'intersection de tous les éléments de $\hat{S}$ qui contiennent $T(A)$

$$\begin{aligned}A &= \{\omega_4, \omega_6\} \\ T(A) &= \{T(\omega_4), T(\omega_6)\} = \{(1, 1, 1, 0), (2, 1, 0, 0)\} \\ \hat{C}(A) &= \{1, 2\} \times \{1\} \times \Omega_Y \times \{0\}.\end{aligned}$$

Plus exactement, nous prolongeons les informations définies sur les pavés de $\hat{S}$ à tous les événements de la tribu $\hat{S}$ de la manière indiquée à la fin du paragraphe II.4 : $I(T(A)) = I(\hat{C}(A))$. Remarquons que dans cet exemple $\hat{C}(A)$ est formé de quatre points de $\hat{\Omega}$ et qu'il contient strictement $T(A)$.

Projection C_F d'un élément $\hat{C} \in \hat{S}$ sur $\prod_F \Omega_\xi$.

Nous supposons que l'ensemble O est fini et nous notons $F = P(O)$.

Soit $F \in F$, $F \neq \emptyset$; nous utilisons les notations suivantes (I.1.3.) :

$$\hat{\Omega} = \prod_F \Omega_\xi \times \prod_{F^c} \Omega_\xi$$

$$\hat{S} = (\otimes_F S_\xi) \otimes (\otimes_{F^c} S_\xi)$$

$C_F \in \otimes_F S_\xi$ projection de $\hat{C} \in \hat{S}$ sur l'espace $\prod_F \Omega_\xi$:

$$C_F = \{x_F \in \prod_F \Omega_\xi \mid (x_F, x_{F^c}) \in \hat{C}\}$$

et nous notons $\hat{C}_F \in \otimes_F S_\xi$ le cylindre de $\hat{\Omega}$ de base C_F : $\hat{C}_F = C_F \times \prod_{F^c} \Omega_\xi$

Exemple : $O = \{\alpha, \beta, \gamma, \delta\}$.

$$F = \{\alpha, \beta\}, \prod_F \Omega_\xi = \Omega_\alpha \times \Omega_\beta = \{0, 1, 2\} \times \{0, 1\}$$

$$\hat{C} = \{0, 1\} \times \Omega_\beta \times \{0\} \times \{1\} \in \hat{S}$$

$$C_F = \{0, 1\} \times \Omega_\beta \in \otimes_F S_\xi$$

$$\hat{C}_F = \{0, 1\} \times \Omega_\beta \times \Omega_\gamma \times \Omega_\delta \in \otimes_F S_\xi$$

Quand nous utilisons des pavés $\hat{C} = \prod_{\xi \in O} C_\xi$ de $\hat{S}$ la projection de $\hat{C}$ sur Ω_ξ (respectivement : $\Omega_{\xi_1} \times \Omega_{\xi_2}$) n'est autre que

la composante C_ξ (resp. $C_{\xi_1} \times C_{\xi_2}$) du pavé $\hat{C}$ dans (Ω_ξ, S_ξ) (resp. $(\Omega_{\xi_1} \times \Omega_{\xi_2}, S_{\xi_1} \otimes S_{\xi_2})$), ce que l'on peut noter

$$C_{\{\xi\}} = C_\xi$$

$$C_{\{\xi_1, \xi_2\}} = C_{\xi_1} \times C_{\xi_2}$$

et plus généralement $C_F = \prod_{\xi \in F} C_\xi$.

VI.4.2. - INTERPRETATION DES RESULTATS DU CHAPITRE II.

Soit I une information quelconque au sens (I.1.1.) sur l'espace $(\hat{\Omega}, \hat{S})$; l'application J définie ci-dessous est une information sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ au sens I.2.2. - I.2.6. avec $F_0 = \{\emptyset\}$.

$$\begin{aligned} \forall (\hat{C}, F) \in \hat{S} \times F \quad J(\hat{C}, F) &= I(\hat{C}_F) \quad \text{si } F \neq \emptyset \\ J(\hat{C}, F) &= 0 \quad \text{si } F = \emptyset \end{aligned}$$

Exemple : $\hat{C} = \{1, 2\} \times \{1\} \times \Omega_\gamma \times \{0\}$

$F_0 = \emptyset$		$J(\hat{C}, \emptyset) = 0$
$F_1 = \{\alpha\}$	$\hat{C}_{F_1} = \{1, 2\} \times \Omega_\beta \times \Omega_\gamma \times \Omega_\delta$	$J(\hat{C}, F_1) = I(\hat{C}_{F_1})$
$F_2 = \{\alpha, \beta\}$	$\hat{C}_{F_2} = \{1, 2\} \times \{1\} \times \Omega_\gamma \times \Omega_\delta$	$J(\hat{C}, F_2) = I(\hat{C}_{F_2})$
$F_3 = \{\alpha, \beta, \gamma\}$	$\hat{C}_{F_3} = \{1, 2\} \times \{1\} \times \Omega_\gamma \times \Omega_\delta$	$J(\hat{C}, F_3) = I(\hat{C}_{F_3})$
$F_4 = 0$	$\hat{C}_0 = \hat{C}$	$J(\hat{C}, 0) = I(\hat{C})$

dans cet exemple, nous avons

$$\emptyset = F_0 \subset F_1 \subset F_2 \subset F_3 \subset F_4 = \Omega$$

et

$$\hat{C}_{F_1} \supset \hat{C}_{F_2} = \hat{C}_{F_3} \supset \hat{C},$$

d'où

$$0 = J(\hat{C}, \emptyset) \leq J(\hat{C}, F_1) \leq J(\hat{C}, F_2) = J(\hat{C}, F_3) \leq J(\hat{C}, \Omega) = I(\hat{C}).$$

Nous avons défini J à partir de l'information I ; il est clair que si l'on connaît J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ on peut retrouver I sur $(\hat{\Omega}, \hat{S})$ en posant :

$$\forall \hat{C} \in \hat{S} \quad I(\hat{C}) = J(\hat{C}, \Omega).$$

Attention : il est important de remarquer qu'étant donné une information J quelconque sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ on peut toujours lui associer une information I sur $(\hat{\Omega}, \hat{S})$ par la relation précédente, mais en général, l'information définie à partir de I comme nous l'avons indiqué au début de ce paragraphe est différente de J (voir II.1 et II.3).

Dans la suite, nous ne nous intéressons qu'aux informations J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ qui sont définies à partir d'une information I sur $(\hat{\Omega}, \hat{S})$ suivant la méthode que nous venons d'indiquer.

VI.4.3. - INFORMATION K SUR $(\hat{\Omega} \times 0, \hat{S} \times F)$

Soit $\{\Pi_\xi | \Pi_\xi \subset S_\xi \quad \xi \in O\}$ une famille finie de partitions finies des ensembles $\{S_\xi | \xi \in O\}$. Nous notons $\hat{\Pi}_\xi$ la partition de $\hat{\Omega}$ formée par les cylindres $\hat{S}$ -mesurables dont les bases sont les éléments de Π_ξ , et $\hat{\Pi}$:

$$\hat{\Pi} = \bigcap_{\xi \in O} \hat{\Pi}_{\xi} = \bigcap_O \hat{\Pi}_{\xi}$$

et aussi pour $F \in \mathcal{F}$, $F \neq \emptyset$

$$\hat{\Pi}_F = \bigcap_{\xi \in F} \hat{\Pi}_{\xi} = \bigcap_F \hat{\Pi}_{\xi} \subset \bigotimes_F S_{\xi}$$

on notera Π_F la partition de $\Pi_F \Omega_{\xi}$ formée par les bases des cylindres de $\hat{\Pi}_F$; il est clair que $\Pi_{\{\xi\}} = \Pi_{\xi}$ et l'on dira parfois que Π_{ξ} (respectivement Π_F) est la composante sur Ω_{ξ} (resp. $\Pi_F \Omega_{\xi}$) de $\hat{\Pi}$. Enfin, nous avons évidemment $\hat{\Pi}_F = \{\hat{C}_F | \hat{C} \in \hat{\Pi}\}$ et $\hat{\Pi} = \hat{\Pi}_F \cap \hat{\Pi}_{F^c}$.

Dans la suite, nous notons $\hat{E}$ l'ensemble des partitions de $\hat{\Omega}$ de la forme de $\hat{\Pi}$.

Etant donné une probabilité P sur $(\hat{\Omega}, \hat{S})$ une information I au sens I.1.1. sur $(\hat{\Omega}, \hat{S})$ et l'information J sur $(\hat{\Omega} \times O, \hat{S} \times F)$ construite à partir de I suivant la méthode indiquée dans le paragraphe VI.4.2. nous avons les résultats suivants :

L'application K définie ci-après sur $\hat{E} \times F$ est une information moyenne apportée par un ensemble d'observateurs sur une partition au sens où nous l'avons définie aux paragraphes VI.3.1. et VI.3.2.

$$\begin{aligned} \forall (\hat{\Pi}, F) \in \hat{E} \times F \\ K(\hat{\Pi}, F) = \sum_{\hat{C} \in \hat{\Pi}} P(\hat{C}) \cdot J(\hat{C}, F). \end{aligned}$$

Nous savons par ailleurs que pour $F \neq \emptyset$

$$H(\hat{\Pi}_F) = \sum_{\hat{C} \in \hat{\Pi}_F} P(\hat{C}) \cdot I(\hat{C})$$

Nous allons montrer que $\forall (\hat{\Pi}, F) \in \hat{E} \times F, F \neq \emptyset :$

$$K(\hat{\Pi}, F) = H(\hat{\Pi}_F).$$

En effet $\hat{C} \in \hat{\Pi} = \bigcap_0 \hat{\Pi}_\xi$ s'écrit de manière unique sous la forme ;
 $\hat{C} = \hat{C}_F \cap \hat{C}_{Fc}$ avec $\hat{C}_F \in \hat{\Pi}_F = \bigcap_F \hat{\Pi}_\xi$ et $\hat{C}_{Fc} \in \hat{\Pi}_{Fc} = \bigcap_{Fc} \hat{\Pi}_\xi$. Nous avons donc
 évidemment

$$\hat{\Pi} = \{ \hat{C} = \hat{C}_1 \cap \hat{C}_2 \mid \hat{C}_1 \in \hat{\Pi}_F, \hat{C}_2 \in \hat{\Pi}_{Fc} \}$$

de plus $J(\hat{C}, F) = I(\hat{C}_F)$ ce qui entraîne

$$K(\hat{\Pi}, F) = \sum_{\substack{\hat{C}_1 \in \hat{\Pi}_F \\ \hat{C}_2 \in \hat{\Pi}_{Fc}}} P(\hat{C}_1 \cap \hat{C}_2) \cdot I(\hat{C}_1)$$

$$K(\hat{\Pi}, F) = \sum_{\hat{C}_1 \in \hat{\Pi}_F} P(\hat{C}_1) \cdot I(\hat{C}_1) = H(\hat{\Pi}_F).$$

Notons que ce résultat est un cas particulier du résultat suivant :
*étant donné une information H sur $\hat{E}$ l'application K définie ci-dessous
 sur $\hat{E} \times F$ est une information au sens du paragraphe VI.3.1.*

$$\forall (\hat{\Pi}, F) \in \hat{E} \times F \quad \text{et} \quad F \neq \emptyset$$

$$K(\hat{\Pi}, F) = H(\hat{\Pi}_F)$$

$$\text{et} \quad K(\hat{\Pi}, \emptyset) = 0.$$

On voit que l'on pourrait reprendre pour H et K des études comparables à celles du paragraphe II.1.

Etant donné $\hat{\Pi} \in \hat{E}, \xi, \xi_1, \xi_2 \in \emptyset$ nous avons donc aussi

$$K(\hat{\Pi}, \{\xi\}) = H(\hat{\Pi}_{\xi})$$

$$K(\hat{\Pi}, \{\xi_1, \xi_2\}) = H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2})$$

où Π_{ξ} , Π_{ξ_1} et Π_{ξ_2} sont les partitions composantes de Π respectivement sur Ω_{ξ} , Ω_{ξ_1} et Ω_{ξ_2} .

VI.5 - DEPENDANCE ET INDEPENDANCE DES OBSERVATEURS.

VI.5.1. - INTERPRETATION DES RESULTATS DU PARAGRAPHE III.1.

$\emptyset$ ensemble fini d'observateurs indépendants.

Nous supposons dans le paragraphe III.1 que les groupes disjoints d'observateurs de $\emptyset$ sont J-indépendants, c'est-à-dire que pour tout

$$(F_1, F_2) \in F \times F \text{ et } F_1 \cap F_2 = \emptyset$$

$$\forall \hat{C} \in \hat{S} \quad J(\hat{C}, (F_1 \cup F_2)) = J(\hat{C}, F_1) + J(\hat{C}, F_2)$$

nous obtenons alors

$$\forall (\hat{C}, F) \in \hat{S} \times F \quad J(\hat{C}, F) = \sum_{\xi \in F} J(\hat{C}, \{\xi\}) = \sum_{\xi \in F} I(\hat{C}_{\xi})$$

Inversement toute information J définie à partir de cette dernière formule est telle que tous les groupes disjoints d'observateurs sont indépendants. Enonçons précisément les résultats que nous employons dans la suite.

Etant donné $\{(\Omega_{\xi}, S_{\xi}, I_{\xi}) | \xi \in \emptyset\}$ une famille finie d'espaces mesurables (finis) (Ω_{ξ}, S_{ξ}) sur lesquels on a défini une information I_{ξ} au sens I.1.1. et une mesure positive, λ sur $(\emptyset, F)$ de masse totale 1 vérifiant : $\xi \in \emptyset$, $\lambda(\xi) > 0$, nous définissons les informations I sur $(\hat{\Omega}, \hat{S})$ et J sur $(\hat{\Omega} \times \emptyset, \hat{S} \times F)$ de la manière suivante :

$\forall \hat{C} \in \hat{S}, \forall F \in F$, en notant $C_\xi = C_{\{\xi\}}$ la projection de $\hat{C}$ sur Ω_ξ et $\hat{C}_\xi$ cylindre associé :

$$I(\hat{C}) = \sum_{\xi \in \mathcal{O}} \lambda(\xi) \cdot I_\xi(C_\xi)$$

$$J(\hat{C}, F) = I(\hat{C}_F) = \sum_{\xi \in F} \lambda(\xi) \cdot I_\xi(C_\xi) \quad \text{si } F \neq \emptyset$$

$$J(\hat{C}, F) = 0 \quad \text{si } F = \emptyset$$

Notons en particulier que : $I(\hat{C}_\xi) = \lambda(\xi) \cdot I_\xi(C_\xi)$.

Par exemple (II.4.1.) : $\hat{C} = \{1,2\} \times \{1\} \times \Omega_Y \times \{0\}$ et $F = \{\alpha, \beta\}$;

supposons que : $\lambda(\alpha) = \lambda(\beta) = \lambda(\gamma) = \frac{2}{7}$, $\lambda(\delta) = \frac{1}{7}$

$$I(\hat{C}) = \frac{2}{7} \cdot I_\alpha(\{1,2\}) + \frac{2}{7} \cdot I_\beta(\{1\}) + \frac{2}{7} \cdot I_\gamma(\Omega_Y) + \frac{1}{7} \cdot I_\delta(\{0\})$$

$$J(\hat{C}, F) = I(\hat{C}_F) = \frac{2}{7} \cdot I_\alpha(\{1,2\}) + \frac{2}{7} \cdot I_\beta(\{1\})$$

$$\text{car} \quad I_\gamma(\Omega_Y) = I_\delta(\Omega_\delta) = 0$$

Si les I_ξ sont définies à partir de la probabilité P sur $(\hat{\Omega}, \hat{S})$ on peut prendre par exemple une des informations étudiées aux paragraphes VI.1.4, ou VI.2.3.

$$I_\xi^*(C_\xi) = \text{Log} \frac{1}{P(\hat{C}_\xi)} \quad \text{ou} \quad I_\xi^0(C_\xi) = \frac{1}{P(\hat{C}_\xi)} - 1$$

en prenant cette deuxième formule on a :

$$J(\hat{C}, F) = \frac{2}{7} \cdot \left(\frac{1}{1/2} - 1 \right) + \frac{2}{7} \cdot \left(\frac{1}{4/6} - 1 \right) = \frac{3}{7}$$

$$I(\hat{C}) = J(\hat{C}, \mathcal{O}) = \frac{4}{7}$$

VI.5.2. - INFORMATION K.

0 ensemble fini d'observateurs indépendants.

Etant donné les informations I sur $(\hat{\Omega}, \hat{S})$ et J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ que nous venons de définir (VI.5.1.), ainsi que P la mesure de probabilité sur $(\hat{\Omega}, \hat{S})$, nous obtenons, avec les notations du paragraphe VI.4.3., les résultats suivants :

Soient une partition $\hat{\Pi} \in \hat{E}$; $\xi, \xi_1, \xi_2 \in 0$ et les partitions $\Pi_{\xi}, \Pi_{\xi_1}, \Pi_{\xi_2}$ composantes de $\hat{\Pi}$ sur $\Omega_{\xi}, \Omega_{\xi_1}, \Omega_{\xi_2}$; nous avons

$$H(\hat{\Pi}_{\xi}) = \sum_{\hat{C}_{\xi} \in \hat{\Pi}_{\xi}} P(\hat{C}_{\xi}) \cdot I(\hat{C}_{\xi}).$$

Nous savons que

$$I(\hat{C}_{\xi}) = \lambda(\xi) \cdot I_{\xi}(C_{\xi})$$

posons

$$H_{\xi}(\Pi_{\xi}) = \sum_{C_{\xi} \in \Pi_{\xi}} P(\hat{C}_{\xi}) \cdot I_{\xi}(C_{\xi})$$

il vient

$$K(\hat{\Pi}, \{\xi\}) = H(\hat{\Pi}_{\xi}) = \lambda(\xi) \cdot H_{\xi}(\Pi_{\xi}).$$

De même

$$H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) = \sum_{(\hat{C}_{\xi_1}, \hat{C}_{\xi_2}) \in \hat{\Pi}_{\xi_1} \times \hat{\Pi}_{\xi_2}} P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2}) \cdot I(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2})$$

or

$$I(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2}) = \lambda(\xi_1) \cdot I_{\xi_1}(C_{\xi_1}) + \lambda(\xi_2) \cdot I_{\xi_2}(C_{\xi_2})$$

ce qui entraîne

$$K(\hat{\Pi}, \{\xi_1, \xi_2\}) = H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) = \lambda(\xi_1) \cdot H_{\xi_1}(\Pi_{\xi_1}) + \lambda(\xi_2) \cdot H_{\xi_2}(\Pi_{\xi_2})$$

c'est-à-dire

$$H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) = H(\hat{\Pi}_{\xi_1}) + H(\hat{\Pi}_{\xi_2}).$$

D'une manière générale, nous aurons pour $F \in \mathcal{F}$ et $F \neq \emptyset$

$$K(\hat{\Pi}, F) = H(\hat{\Pi}_F) = \sum_{\xi \in F} \lambda(\xi) \cdot H_{\xi}(\Pi_{\xi})$$

$$H(\hat{\Pi}_F) = \sum_{\xi \in F} H(\hat{\Pi}_{\xi}).$$

On trouvera un exemple d'information de ce type au paragraphe VII.1.

Dans le paragraphe VI.5.1. nous sommes partis de l'hypothèse que les groupes disjoints d'observateurs de $(\mathcal{O}, F)$ sont J -indépendants et comme prévu, dans le paragraphe VI.3.2., nous obtenons bien que les groupes disjoints d'observateurs de $(\mathcal{O}, F)$ sont K -indépendants ; en effet

$$(F_1, F_2) \in F \times F \text{ et } F_1 \cap F_2 = \emptyset \implies$$

$$K(\hat{\Pi}, F_1 \cup F_2) = K(\hat{\Pi}, F_1) + K(\hat{\Pi}, F_2)$$

c'est-à-dire

$$K(\hat{\Pi}, (F_1, F_2)) = 0.$$

VI.5.3. - INTERPRETATION DES RESULTATS DU PARAGRAPHE III.2.

0 ensemble fini d'observateurs non-indépendants.

Supposer seulement que les groupes d'observateurs de (O, F) sont non-indépendants ne donne aucun moyen de déterminer une forme générale pour I sur $(\hat{\Omega}, \hat{S})$ ou J sur $(\hat{\Omega} \times O, \hat{S} \times F)$; aussi avons nous été amenés à poser dans le paragraphe III.2. d'autres hypothèses concernant la dépendance et l'indépendance des observateurs.

Enonçons précisément le résultat que nous utilisons par la suite.

Soient O_1 et O_2 deux ensembles finis, non vides, $O = O_1 \cup O_2$, $O_1 \cap O_2 = \emptyset$, $F_1 = P(O_1)$, $F_2 = P(O_2)$, $F = P(O)$; la famille $\{(\Omega_\xi, S_\xi, I_\xi) | \xi \in O\}$ dont chaque élément est un espace mesurable fini (Ω_ξ, S_ξ) sur lequel on a défini une information I_ξ au sens I.1.1. ; et enfin une famille $\{(\Omega_{\xi_1} \times \Omega_{\xi_2}, S_{\xi_1} \otimes S_{\xi_2}, I_{\xi_1 \xi_2}) | (\xi_1, \xi_2) \in O_1 \times O_2\}$ dont chaque élément est un produit de deux espaces mesurables finis (de la famille précédente) sur lequel on a défini une information $I_{\xi_1 \xi_2}$ au sens I.1.1.

Nous supposons que $\forall \xi_1 \in O_1, \forall \xi_2 \in O_2, \forall (C_{\xi_1}, C_{\xi_2}) \in S_{\xi_1} \times S_{\xi_2}$

on a $I_{\xi_1}(C_{\xi_1}) = I_{\xi_1 \xi_2}(C_{\xi_1} \times \Omega_{\xi_2})$; $I_{\xi_2}(C_{\xi_2}) = I_{\xi_1 \xi_2}(\Omega_{\xi_1} \times C_{\xi_2})$

Soit λ une mesure positive finie de masse totale égale à 1 sur $(O_1 \times O_2, F_1 \otimes F_2)$. Nous écrivons $\lambda(\xi_1, \xi_2)$ la probabilité du couple $(\xi_1, \xi_2) \in O_1 \times O_2$ et nous utilisons les notations condensées suivantes pour

$$F_i \in F_i, \xi_i \in O_i, i = 1, 2 : \lambda(F_1, F_2) = \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2)$$

$$\lambda_1(\xi_1) = \lambda(\xi_1, O_2) ; \lambda_2(\xi_2) = \lambda(O_1, \xi_2) .$$

Supposons aussi que les probabilités marginales sont strictement positives :

$$\xi_i \in O_i \implies \lambda_i(\xi_i) > 0 \quad i = 1, 2.$$

Avec ces notations, nous définissons I sur $(\hat{\Omega}, \hat{S}) = (\prod_{\xi} \Omega_{\xi}, \otimes_{\xi} S_{\xi})$ et J sur $(\hat{\Omega} \times O, \hat{S} \times F)$ de la manière suivante :

$$\forall \hat{C} \in \hat{S}, \forall F_1 \in F_1, \forall F_2 \in F_2, F = F_1 \cup F_2 \neq \emptyset$$

$$I(\hat{C}) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(C_{\{\xi_1, \xi_2\}})$$

en notant $C_{\{\xi_1, \xi_2\}}$ la projection de $\hat{C}$ sur $\Omega_{\xi_1} \times \Omega_{\xi_2}$, et $C_{\xi_i} = C_{\{\xi_i\}}$, $i = 1, 2$

les projections de $\hat{C}$ respectivement sur Ω_{ξ_1} et Ω_{ξ_2} , il vient :

$$\begin{aligned} J(\hat{C}, F) = I(\hat{C}_F) &= \sum_{\xi_1 \in F_1} \lambda(\xi_1, O_2 - F_2) \cdot I_{\xi_1}(C_{\xi_1}) \\ &+ \sum_{\xi_2 \in F_2} \lambda(O_1 - F_1, \xi_2) \cdot I_{\xi_2}(C_{\xi_2}) \\ &+ \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(C_{\{\xi_1, \xi_2\}}) \end{aligned}$$

ce que l'on peut écrire aussi, si $\hat{C}$ est un pavé mesurable de $\otimes_{\xi} S_{\xi}$

$$\begin{aligned} J(\hat{C}, F) = I(\hat{C}_F) &= \sum_{\xi_1 \in F_1} \lambda_1(\xi_1) \cdot I_{\xi_1}(C_{\xi_1}) \\ &+ \sum_{\xi_2 \in F_2} \lambda_2(\xi_2) \cdot I_{\xi_2}(C_{\xi_2}) - \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(C_{\xi_1}; C_{\xi_2}) \end{aligned}$$

avec $I_{\xi_1 \xi_2}(C_{\xi_1}; C_{\xi_2}) = I_{\xi_1}(C_{\xi_1}) + I_{\xi_2}(C_{\xi_2}) - I_{\xi_1 \xi_2}(C_{\xi_1} \times C_{\xi_2})$,

car dans ce cas $C_{\xi_1} \times C_{\xi_2} = C_{\{\xi_1, \xi_2\}}$. Pour simplifier l'écriture nous notons

$I_{\xi_1 \xi_2}(C_{\xi_1}; C_{\xi_2})$ à la place de $I_{\xi_1 \xi_2}((C_{\xi_1} \times \Omega_{\xi_2}); (\Omega_{\xi_1} \times C_{\xi_2}))$.

Ces informations I et J vérifient les hypothèses que nous avons évoquées au début de ce paragraphe. Notons aussi que, pour $\hat{C} \in \hat{S}$ donné, on a

$$\forall \xi \in O \quad I(\hat{C}_\xi) = \lambda(\xi) \cdot I_\xi(C_\xi)$$

$$\forall (\xi_1, \xi'_1) \in O_1 \times O_1, \xi_1 \neq \xi'_1, i = 1, 2$$

$$I(\hat{C}_{\xi_1} \cap \hat{C}_{\xi'_1}) = \lambda(\xi_1) \cdot I_{\xi_1}(C_{\xi_1}) + \lambda(\xi'_1) \cdot I_{\xi'_1}(C_{\xi'_1}) \quad i = 1, 2$$

$$\forall F_i \in F_i, F_i \neq \emptyset, I(\hat{C}_{F_i}) = \sum_{\xi \in F_i} \lambda(\xi) \cdot I_\xi(C_\xi) \quad i = 1, 2$$

enfin pour un pavé mesurable $\hat{C}$ de $\hat{S}$ on a aussi

$$F_1 \in F_1, F_2 \in F_2 \text{ et } F = F_1 \cup F_2 \in F$$

$$I(\hat{C}_F) = I(\hat{C}_{F_1}) + I(\hat{C}_{F_2}) - I(\hat{C}_{F_1}; \hat{C}_{F_2})$$

c'est-à-dire encore

$$J(\hat{C}, F) = J(\hat{C}, F_1) + J(\hat{C}, F_2) - J((\hat{C}, F_1); (\hat{C}, F_2))$$

$$\text{avec } I(\hat{C}_{F_1}; \hat{C}_{F_2}) = J((\hat{C}, F_1); (\hat{C}, F_2)) = \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot I_{\xi_1 \xi_2}(C_{\xi_1}; C_{\xi_2}).$$

En particulier pour $O = O_1 \cup O_2$ et $\hat{C}$ pavé mesurable de $\hat{S}$ nous avons :

$$J(\hat{C}, 0) = J(\hat{C}, 0_1) + J(\hat{C}, 0_2) - J((\hat{C}, 0_1), (\hat{C}, 0_2))$$

avec

$$J((\hat{C}, 0_1), (\hat{C}, 0_2)) = \sum_{(\varepsilon_1, \varepsilon_2) \in 0_1 \times 0_2} \lambda(\varepsilon_1, \varepsilon_2) \cdot I_{\varepsilon_1 \varepsilon_2}(C_{\varepsilon_1}, C_{\varepsilon_2})$$

Reprenons l'exemple du paragraphe VI.4.1. avec

$$0_1 = \{\alpha, \beta\}, \quad 0_2 = \{\gamma, \delta\}, \quad 0 = 0_1 \cup 0_2.$$

Supposons que le tableau suivant nous définisse la pondération λ des observateurs sur $(0_1 \times 0_2, F_1 \otimes F_2)$

Pondération λ	γ	δ	Total
α	0,4	0,2	0,6
β	0,1	0,3	0,4
Total	0,5	0,5	1

$$\text{Soit } \hat{C} = \{1, 2\} \times \{1\} \times \Omega_Y \times \{0\}$$

$$F_1 = 0_1 = \{\alpha, \beta\}$$

$$F_2 = \{\delta\}$$

$$F = F_1 \cup F_2 = \{\alpha, \beta, \delta\}.$$

Remarquons que $\hat{C} = \hat{C}_F$.

Avec des notations évidentes, il vient :

$$\begin{aligned} J(\hat{C}, O_1) &= 0,6 \cdot I_{\alpha}(\{1,2\}) + 0,4 \cdot I_{\beta}(\{1\}) \\ J(\hat{C}, O_2) &= J(\hat{C}, F_2) = 0,5 \cdot I_{\delta}(\{0\}) \\ J(\hat{C}, F) &= 0,4 \cdot I_{\alpha}(\{1,2\}) + 0,1 \cdot I_{\beta}(\{1\}) \\ &\quad + 0,2 \cdot I_{\alpha\delta}(\{1,2\} \times \{0\}) + 0,3 \cdot I_{\beta\delta}(\{1\} \times \{0\}) \\ I(\hat{C}) &= 0,4 \cdot I_{\alpha\gamma}(\{1,2\} \times \Omega_{\gamma}) + 0,2 \cdot I_{\alpha\delta}(\{1,2\} \times \{0\}) \\ &\quad + 0,1 \cdot I_{\beta\gamma}(\{1\} \times \Omega_{\gamma}) + 0,3 \cdot I_{\beta\delta}(\{1\} \times \{0\}), \end{aligned}$$

comme on sait que :

$$\begin{aligned} I_{\alpha\gamma}(\{1,2\} \times \Omega_{\gamma}) &= I_{\alpha}(\{1,2\}) \\ I_{\beta\gamma}(\{1\} \times \Omega_{\gamma}) &= I_{\beta}(\{1\}). \end{aligned}$$

On retrouve bien que :

$$J(\hat{C}, F) = I(\hat{C}_F) = I(\hat{C}).$$

Supposons que les informations $\{I_{\xi_1 \xi_2} | \xi_1 \in O_1, \xi_2 \in O_2\}$ sont définies à partir de la probabilité P sur $(\hat{\Omega}, \hat{S})$; on peut prendre, par exemple, une des informations étudiées aux paragraphes VI.1.4. et VI.2.3.

$$I_{\xi_1 \xi_2}^*(C_{\xi_1} \times C_{\xi_2}) = \text{Log} \frac{1}{P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2})}$$

$$\text{ou } I_{\xi_1 \xi_2}^0(C_{\xi_1} \times C_{\xi_2}) = \frac{1}{P(\hat{C}_{\xi_1})} + \frac{1}{P(\hat{C}_{\xi_2})} - \frac{P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2})}{P(\hat{C}_{\xi_1}) \cdot P(\hat{C}_{\xi_2})} - 1$$

et avec cette dernière formule, il vient :

$$J(\hat{C}, 0_1) = 0,8$$

$$J(\hat{C}, 0_2) = J(\hat{C}, F_2) = 0,5$$

$$J(\hat{C}) = J(\hat{C}, F) = 1,233.$$

VI.5.4. - INFORMATION K.

0 ensemble fini d'observateurs non indépendants.

Avec les notations du paragraphe VI.4.3. étant donné les informations I sur $(\hat{\Omega}, \hat{S})$ et J sur $(\hat{\Omega} \times 0, \hat{S} \times F)$ que nous venons de définir (VI.5.3.) ainsi que P la mesure de probabilité sur $(\hat{\Omega}, \hat{S})$ nous obtenons les résultats suivants :

Soient une partition $\hat{\Pi} \in \hat{E}, \xi, \xi_1, \xi_2 \in 0$ et les partitions $\Pi_\xi, \Pi_{\xi_1}, \Pi_{\xi_2}$ composantes respectives de Π sur $\Omega_\xi, \Omega_{\xi_1}, \Omega_{\xi_2}$; nous avons :

$$H(\hat{\Pi}_\xi) = \sum_{\hat{C}_\xi \in \hat{\Pi}_\xi} P(\hat{C}_\xi) \cdot I(\hat{C}_\xi)$$

nous savons que

$$I(\hat{C}_\xi) = \lambda(\xi) \cdot I_\xi(C_\xi)$$

il vient

$$K(\hat{\Pi}, \{\xi\}) = H(\hat{\Pi}_\xi) = \lambda(\xi) \cdot H_\xi(\Pi_\xi)$$

en notant

$$H_\xi(\Pi_\xi) = \sum_{C_\xi \in \Pi_\xi} P(\hat{C}_\xi) \cdot I_\xi(C_\xi).$$

De même si $\xi_1, \xi_2 \in 0_1, i = 1, 2$ on obtient :

$$K(\hat{\Pi}, \{\xi_1, \xi_2\}) = H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) = \lambda(\xi_1) \cdot H_{\xi_1}(\Pi_{\xi_1}) + \lambda(\xi_2) \cdot H_{\xi_2}(\Pi_{\xi_2})$$

c'est-à-dire

$$H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) = H(\hat{\Pi}_{\xi_1}) + H(\hat{\Pi}_{\xi_2})$$

par contre si $(\xi_1, \xi_2) \in \mathcal{O}_1 \times \mathcal{O}_2$ il vient :

$$\begin{aligned} K(\hat{\Pi}, \{\xi_1, \xi_2\}) &= H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) \\ &= \lambda(\xi_1, (\mathcal{O}_2 - \{\xi_2\})) \cdot H_{\xi_1}(\Pi_{\xi_1}) \\ &\quad + \lambda((\mathcal{O}_1 - \{\xi_1\}), \xi_2) \cdot H_{\xi_2}(\Pi_{\xi_2}) \\ &\quad + \lambda(\xi_1, \xi_2) \cdot H_{\xi_1 \xi_2}(\Pi_{\xi_1} \times \Pi_{\xi_2}) \end{aligned}$$

$$\text{avec } H_{\xi_1 \xi_2}(\Pi_{\xi_1} \times \Pi_{\xi_2}) = \sum_{(C_{\xi_1}, C_{\xi_2}) \in \Pi_{\xi_1} \times \Pi_{\xi_2}} P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2}) \cdot I_{\xi_1 \xi_2}(C_{\xi_1} \times C_{\xi_2})$$

en notant conventionnellement $\Pi_{\xi_1} \times \Pi_{\xi_2}$ la partition de $\Omega_{\xi_1} \times \Omega_{\xi_2}$ formée de l'ensemble des pavés $\{C_{\xi_1} \times C_{\xi_2} \mid (C_{\xi_1}, C_{\xi_2}) \in \Pi_{\xi_1} \times \Pi_{\xi_2}\}$. On peut écrire aussi

$$H(\hat{\Pi}_{\xi_1} \cap \hat{\Pi}_{\xi_2}) = H(\hat{\Pi}_{\xi_1}) + H(\hat{\Pi}_{\xi_2}) - H(\hat{\Pi}_{\xi_1}; \hat{\Pi}_{\xi_2})$$

avec

$$H(\hat{\Pi}_{\xi_1}; \hat{\Pi}_{\xi_2}) = \lambda(\xi_1, \xi_2) \cdot H_{\xi_1 \xi_2}(\Pi_{\xi_1}; \Pi_{\xi_2})$$

et

$$H_{\xi_1 \xi_2}(\Pi_{\xi_1}; \Pi_{\xi_2}) = \sum_{(C_{\xi_1}, C_{\xi_2}) \in \Pi_{\xi_1} \times \Pi_{\xi_2}} P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2}) \cdot I_{\xi_1 \xi_2}(C_{\xi_1}; C_{\xi_2})$$

D'une manière générale pour $F_1 \in \mathcal{F}_1$ et $F_2 \in \mathcal{F}_2$, nous aurons

$$K(\hat{\Pi}, F_i) = H(\hat{\Pi}_{F_i}) = \sum_{\xi \in F_i} \lambda(\xi) \cdot H_{\xi}(\pi_{\xi}) \quad i = 1, 2$$

et pour $F = F_1 \cup F_2$

$$K(\hat{\Pi}, F) = H(\hat{\Pi}_F) = H(\hat{\Pi}_{F_1}) + H(\hat{\Pi}_{F_2}) - H(\hat{\Pi}_{F_1}; \hat{\Pi}_{F_2})$$

avec

$$K(\hat{\Pi}, (F_1; F_2)) = H(\hat{\Pi}_{F_1}; \hat{\Pi}_{F_2}) = \sum_{(\xi_1, \xi_2) \in F_1 \times F_2} \lambda(\xi_1, \xi_2) \cdot H_{\xi_1 \xi_2}(\pi_{\xi_1}; \pi_{\xi_2})$$

Et pour

$$O = O_1 \cup O_2.$$

$$K(\hat{\Pi}, O_i) = H(\hat{\Pi}_{O_i}) = \sum_{\xi \in O_i} \lambda(\xi) \cdot H_{\xi}(\pi_{\xi}) \quad i = 1, 2$$

$$K(\hat{\Pi}, O) = K(\hat{\Pi}, O_1) + K(\hat{\Pi}, O_2) - K(\hat{\Pi}, (O_1; O_2))$$

avec

$$K(\hat{\Pi}, (O_1; O_2)) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot H_{\xi_1 \xi_2}(\pi_{\xi_1}; \pi_{\xi_2})$$

VI.5.5. - INFORMATION MUTUELLE $K(\hat{\Pi}, (O_1; O_2))$.

En pratique pour calculer les informations J et K des paragraphes VI.5.3. et VI.5.4. nous allons prendre pour chaque couple $(\xi_1, \xi_2) \in O_1 \times O_2$ des informations $I_{\xi_1 \xi_2}$ et $H_{\xi_1 \xi_2}$ sur $(\Omega_{\xi_1} \times \Omega_{\xi_2}, S_{\xi_1} \otimes S_{\xi_2})$ du type de celles qui sont définies aux paragraphes VI.1.4. et VI.2.4.

Nous avons montré que sous les hypothèses et avec les notations des paragraphes VI.5.3. et VI.5.4. nous avons pour $O = O_1 \cup O_2$

$$\forall \hat{\Pi} \in \hat{E}$$

$$K(\hat{\Pi}, O_i) = \sum_{\xi \in O_i} \lambda(\xi) \cdot H_{\xi}(\Pi_{\xi}) \quad i = 1, 2$$

et

$$K(\hat{\Pi}, O) = K(\hat{\Pi}, O_1) + K(\hat{\Pi}, O_2) - K(\hat{\Pi}, (O_1, O_2))$$

avec

$$K(\hat{\Pi}, (O_1, O_2)) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot H_{\xi_1 \xi_2}(\Pi_{\xi_1}, \Pi_{\xi_2}).$$

Remplaçons dans cette formule $H_{\xi_1 \xi_2}(\Pi_{\xi_1}, \Pi_{\xi_2})$ par l'une des quantités définies aux paragraphes VI.1.4. ou VI.2.4. :

$$H_{\xi_1 \xi_2}^i(\Pi_{\xi_1}, \Pi_{\xi_2}) = \sum_{(C_{\xi_1}, C_{\xi_2}) \in \Pi_{\xi_1} \times \Pi_{\xi_2}} P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2}) \cdot I_{\xi_1 \xi_2}^i(C_{\xi_1}, C_{\xi_2})$$

$$i = 0, 1$$

avec

$$I_{\xi_1 \xi_2}^0(C_{\xi_1}, C_{\xi_2}) = \frac{P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2})}{P(\hat{C}_{\xi_1}) \cdot P(\hat{C}_{\xi_2})} - 1$$

ou

$$I_{\xi_1 \xi_2}^1(C_{\xi_1}, C_{\xi_2}) = I_{\xi_1 \xi_2}^{\star}(C_{\xi_1}, C_{\xi_2}) = \text{Log} \frac{P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2})}{P(\hat{C}_{\xi_1}) \cdot P(\hat{C}_{\xi_2})}.$$

Il vient :

$$K^i(\Pi, (O_1, O_2)) = \sum_{(\xi_1, \xi_2) \in O_1 \times O_2} \lambda(\xi_1, \xi_2) \cdot H_{\xi_1 \xi_2}^i(\Pi_{\xi_1}; \Pi_{\xi_2})$$

$$i = 0, 1$$

Ces deux informations mutuelles possèdent les propriétés que nous avons annoncées au paragraphe VI.3.3. et qui généralisent les propriétés de l'information mutuelle au sens de SHANNON (VI.1.5.).

Avec des notations évidentes, pour $i = 0, 1$:

$K^i(\hat{\Pi}, (O_1; O_2)) = H^i(\hat{\Pi}_{O_1}; \hat{\Pi}_{O_2})$; étant donné la partition $\hat{\Pi}_{O_1}$ fixé les applications $H^i(\hat{\Pi}_{O_1}; \cdot)$ sont des informations au sens du paragraphe VI.1.2. ; en particulier, pour toutes les partitions $\hat{\Pi}_{O_2}, \hat{\Pi}'_{O_2}$ telles que $\hat{\Pi}'_{O_2} \leq \hat{\Pi}_{O_2}$ nous avons

$$0 \leq H^i(\hat{\Pi}_{O_1}; \hat{\Pi}_{O_2}) \leq H^i(\hat{\Pi}_{O_1}; \hat{\Pi}'_{O_2}) \leq H^i(\hat{\Pi}_{O_1})$$

- avec égalité à zéro si et seulement si pour tous les couples $(\xi_1, \xi_2) \in O_1 \times O_2$ tels que $\lambda(\xi_1, \xi_2) > 0$, les partitions Π_{ξ_1}, Π_{ξ_2} sont $H_{\xi_1 \xi_2}^i$ -indépendantes en information, c'est-à-dire que les éléments $C_{\xi_1} \in \Pi_{\xi_1}$ et $C_{\xi_2} \in \Pi_{\xi_2}$ sont $I_{\xi_1 \xi_2}^i$ -indépendants en information,

- avec égalité à l'information maximum $H^i(\Pi_{O_1})$ si et seulement si pour tous les couples $(\xi_1, \xi_2) \in O_1 \times O_2$, tels que $\lambda(\xi_1, \xi_2) > 0$ et pour tout $C_{\xi_2} \in \Pi_{\xi_2}$ il existe $C_{\xi_1} \in \Pi_{\xi_1}$ vérifiant $P(\hat{C}_{\xi_1} \cap \hat{C}_{\xi_2}) = P(\hat{C}_{\xi_2})$.

Dans la pratique, nous pourrions par exemple supposer que O_1 est donné et que $H^1(\Pi, O_1) = K^1(\Pi, O_1)$ représente l'information de référence.

Les quantités $K^1(\Pi', (O_1; O_2'), K^1(\Pi'', (O_1; O_2'')) , \dots$, nous permettront de mesurer et de comparer la manière dont différents groupes d'observateurs $O_2', O_2'', O_2''' , \dots$ résument cette information de référence.

L'exemple traité au début du chapitre suivant nous permettra de préciser davantage encore notre raisonnement.

CHAPITRE VII

EXEMPLES ET APPLICATIONS.

Nous avons traité au paragraphe V.4 un exemple qui nous paraît fondamental pour la théorie : d'une part, il a permis de mieux distinguer les notions d'information dans un système statique et d'information dans un système dynamique, de dégager les liens entre ces deux notions ; d'autre part, cette étude ne fait intervenir, ni la notion d'information moyenne apportée par les éléments d'une partition, ni même la notion de probabilité.

L'exemple que nous allons étudier en détail dans le premier paragraphe de ce chapitre est écrit dans une optique différente; nous utilisons des fréquences et des probabilités afin d'illustrer les résultats du chapitre VI et de montrer clairement comment ces résultats englobent et généralisent ceux de la théorie classique de l'information de SHANNON. Cet exemple nous servira aussi de support pour introduire différentes perspectives d'applications.

VII.1 - EXEMPLE.

Etudions les tableaux de données de la page VII.2. Les notations que nous employons sont celles du chapitre VI avec quelques modifications évidentes :

$$\Omega = \{\omega_1, \dots, \omega_{16}\}, \quad S = P(\Omega).$$

La probabilité P sur l'espace mesurable (Ω, S) est engendrée par :

$$\forall \omega \in \Omega \quad P(\omega) = \frac{1}{16}.$$

TABLEAUX DE DONNEES DE L'EXEMPLE DU PARAGRAPHE VII.1

O_1
 $\alpha_1 \alpha_2 \beta_1 \beta_2 \beta_3 \beta_4 \gamma_1 \gamma_2 \gamma_3 \gamma_4$

O_2
 $\alpha \beta \gamma$

O_3

O_4

O_5

Ω	ω_1	1	1	1	1	1	1	1	1	1	1	1	1			1		1		1	
	ω_2	1	1	1	1	1	1	1	1	1	1	1	1			1		1		1	
	ω_3	1	1	1	1		1	1	1	1	1	1	1			1		1		1	
	ω_4	1	1	1	1	1					1	1				2		1		4	
	ω_5	1	1	1	1	1	1		1		1	1				2		1		4	
	ω_6	1						1	1	1	1	1				3		2		2	
	ω_7	1	1		1						1					4		2		2	
	ω_8	1	1							1	1					4		2		2	
	ω_9	1	1								1					4		2		2	
	ω_{10}		1					1	1	1	1		1				5		3		3
	ω_{11}				1				1	1	1		1				5		3		3
	ω_{12}							1	1	1	1		1				5		3		3
	ω_{13}							1	1	1	1		1				5		3		3
	ω_{14}					1	1			1	1		1				5		3		3
	ω_{15}										1						6		4		1
	ω_{16}																6		4		1

λ_1 : 1 1 1 1 1 1 1 1 1 1 1

λ_2 : 2 4 4

10

10

10

$$O_1 \left[\begin{array}{c} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{array} \right] \left[\begin{array}{c} 1 \\ 1 \\ \\ 1 \\ 1 \\ 1 \\ \\ 1 \\ \\ 1 \end{array} \right] \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right] \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right] \left[\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right]$$

Dans les tableaux, les cases vides sont des zéros.

Les tableaux λ de pondération des observateurs sont donnés en dixièmes.



Nous notons pour tout $\xi \in O_i$, $i = 1, \dots, 5$: $\Omega_\xi = \{0,1\}$ pour O_1 et O_2 ; $\Omega_\xi = \{1, \dots, 6\}$ pour O_3 ; $\Omega_\xi = \{1, \dots, 4\}$ pour O_4 et O_5 ;
 $S_\xi = P(\Omega_\xi)$,

et

$$\hat{\Omega}_i = \prod_{O_i} \Omega_\xi, \quad \hat{S}_i = \bigotimes_{\xi \in O_i} S_\xi \quad i = 1, \dots, 5$$

Soient T_i , $i = 1, \dots, 5$, les applications mesurables de (Ω, S) respectivement dans $(\hat{\Omega}_i, \hat{S}_i)$, $i = 1, \dots, 5$, qui sont définies par les tableaux de la page VII.2. Nous notons P_i les probabilités sur $(\hat{\Omega}_i, \hat{S}_i)$ engendrées par T_i :

$$\forall \hat{C} \in \hat{S}_i \quad P_i(\hat{C}) = P(T_i^{-1}(\hat{C})) \quad i = 1, \dots, 5.$$

Nous notons dans la suite $(\hat{\Omega}, \hat{S})$ l'espace mesurable produit défini par

$$\hat{\Omega} = \hat{\Omega}_1 \times \hat{\Omega}_2 = \prod_{O_1 \cup O_2} \Omega_\xi$$

$$\hat{S} = \hat{S}_1 \otimes \hat{S}_2 = \bigotimes_{O_1 \cup O_2} S_\xi$$

et $\hat{\Pi}$ (resp. $\hat{\Pi}_i$, $i = 1, \dots, 5$) la partition la plus fine de $\hat{\Omega}$ (resp. $\hat{\Omega}_i$, $i = 1, \dots, 5$) c'est-à-dire la partition engendrée par le produit des partitions Π_ξ de Ω_ξ , $\xi \in O_1 \cup O_2$ (resp. $\xi \in O_i$, $i = 1, \dots, 5$) et où chaque Π_ξ est de la forme

$$\Pi_\xi = \{ \{x_\xi\} \mid x_\xi \in \Omega_\xi \}.$$

Nous noterons enfin Π_i , $i = 1, \dots, 5$ les partitions de (Ω, S) de la forme

$$\Pi_i = T_i^{-1}(\hat{\Pi}_{0_i}) \quad i = 1, \dots, 5$$

et $\Pi = \Pi_1 \cap \Pi_2 = T^{-1}(\hat{\Pi})$ où T est l'application (T_1, T_2) de (Ω, S) dans $(\hat{\Omega}, \hat{S})$.

VII.1.1 - INFORMATION SUR UN EVENEMENT.

Nous avons souligné dans l'introduction de ce mémoire que "si l'on attache à l'ensemble Ω une signification concrète que l'on ne désire pas modifier, une meilleure caractérisation d'un élément $\omega \in \Omega$ peut ne pas être associée à une meilleure localisation de cet élément dans la population réellement étudiée"; c'est en particulier pour cette raison que nous avons introduit la notion d'information apportée sur un événement par un groupe d'observateurs.

L'étude suivante indique clairement la solution qu'il faut apporter aux exemples 1 et 2 que nous avons utilisés dans l'introduction pour illustrer cette remarque.

Considérons le tableau $\Omega \times O_1$ de la page VII.2 et posons-nous les questions suivantes : comment déterminer les informations apportées quand on énonce successivement les propositions :

$p = "$ ω répond 1 à l'observateur α_1 $"$

$q = "$ ω répond 0 à l'observateur α_2 $"$

$r = "$ ω répond 1 à l'observateur β_1 $"$

$s = "$ ω répond 1 à l'observateur β_2 $"$

etc...

VII.1.1 - a) Voyons d'abord ce que donne le calcul de l'information classique au sens de SHANNON, notée I^* dans la suite. Définissons pour tout $\xi \in O_1$ l'ensemble A_ξ des éléments de Ω qui répondent 1 à cet observateur et $\hat{C}_\xi$ le cylindre de $(\hat{\Omega}_1, \hat{S}_1)$ de la forme $\hat{C}_\xi = \{1\} \times \prod_{O_1 - \{\xi\}} \Omega_\xi$. Par définition

$A_\xi = T_1^{-1}(\hat{C}_\xi) = \{\omega \in \Omega \mid \omega \text{ répond 1 à l'observateur } \xi\}$. Plaçons nous d'abord dans (Ω, S, P) ; nous avons :

.information apportée par la proposition p

$$I^*(A_{\alpha_1}) = \log_2 \frac{1}{P(A_{\alpha_1})} = 0,830$$

.information apportée par la proposition $[p \text{ et } q]$

$$I^*(A_{\alpha_1} \cap A_{\alpha_2}^c) = \log_2 \frac{1}{P(A_{\alpha_1} \cap A_{\alpha_2}^c)} = 3$$

.information apportée par la proposition $[p \text{ et } q \text{ et } r]$

$$I^*(A_{\alpha_1} \cap A_{\alpha_2}^c \cap A_{\beta_1}) = 4$$

.information apportée par la proposition $[p \text{ et } q \text{ et } r \text{ et } s]$

$$I^*(A_{\alpha_1} \cap A_{\alpha_2}^c \cap A_{\beta_1} \cap A_{\beta_2}) = 4$$

comme $A_{\alpha_1} \cap A_{\alpha_2}^c \cap A_{\beta_1} = A_{\alpha_1} \cap A_{\alpha_2}^c \cap A_{\beta_1} \cap A_{\beta_2} = \{\omega_1\}$ nous obtenons donc

évidemment pour ces deux ensembles la même information. C'est-à-dire que la proposition s qui nous apporte effectivement des renseignements sur ω_1 ne nous permet pas de "mieux" localiser ω_1 dans Ω quand on sait déjà $[p \text{ et } q \text{ et } r]$.

Si l'on se place dans $(\hat{\Omega}_1, \hat{S}_1, P_1)$, la proposition s permet bien une "meilleure" localisation :

$$\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1} \supset_{\text{strict}} \hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1} \cap \hat{C}_{\beta_2}.$$

mais ces événements ont la même probabilité et ont donc aussi la même information. Nous avons en effet :

$$T_1^{-1}(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1}) = T_1^{-1}(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1} \cap \hat{C}_{\beta_2}) = \{\omega_1\}$$

donc

$$I_1^*(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1}) = I_1^*(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1} \cap \hat{C}_{\beta_2}) = 4$$

D'une manière générale pour quantifier l'information apportée par une proposition du type $[p \text{ et } q \text{ et } r]$ on peut se placer, pour l'information au sens de SHANNON, indifféremment sur (Ω, S, P) ou $(\hat{\Omega}_1, \hat{S}_1, P_1)$ car pour tout $\hat{C} \in \hat{S}$ on a $P_1(\hat{C}) = P(T_1^{-1}(\hat{C}))$ ce qui implique :

$$I_1^*(\hat{C}) = I^*(T_1^{-1}(\hat{C})).$$

VII.1.1 - b)

Pour illustrer les résultats des chapitres précédents, nous nous plaçons dans $(\hat{\Omega}_1, \hat{S}_1)$ (on peut, le cas échéant, passer de l'espace $(\hat{\Omega}_1, \hat{S}_1)$ à l'espace (Ω, S) en utilisant les techniques indiquées au paragraphe II.2.)

Supposons que les observateurs de O_1 sont indépendants : prenons les notations et les hypothèses du paragraphe VI.5.1., où la pondération λ des observateurs de O_1 est celle qui est définie dans le tableau λ_1 de

la page VII.2. ; il vient (indice 0 ou 1 selon que l'on utilise sur (Ω_ξ, S_ξ) , $\xi \in O_1$, l'une ou l'autre des informations définies aux paragraphes VI.2.3. ou VI.1.4.) :

.information apportée par la proposition p

$$I^0(\hat{C}_{\alpha_1}) = \lambda_1(\alpha_1) \cdot I_{\alpha_1}^0(\hat{C}_{\alpha_1}) = \lambda_1(\alpha_1) \cdot \left(\frac{1}{P_1(\hat{C}_{\alpha_1})} - 1 \right) = 0,077$$

$$I^1(\hat{C}_{\alpha_1}) = \lambda_1(\alpha_1) \cdot I_{\alpha_1}^*(\hat{C}_{\alpha_1}) = \lambda_1(\alpha_1) \cdot \text{Log}_2 \frac{1}{P_1(\hat{C}_{\alpha_1})} = 0,083$$

.information apportée par la proposition $[p \text{ et } q]$

$$I^0(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c) = \lambda_1(\alpha_1) \cdot I_{\alpha_1}^0(\hat{C}_{\alpha_1}) + \lambda_1(\alpha_2) \cdot I_{\alpha_2}^0(\hat{C}_{\alpha_2}^c) = 0,177$$

$$I^1(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c) = \lambda_1(\alpha_1) \cdot I_{\alpha_1}^*(\hat{C}_{\alpha_1}) + \lambda_1(\alpha_2) \cdot I_{\alpha_2}^*(\hat{C}_{\alpha_2}^c) = 0,183$$

.information apportée par la proposition $[p \text{ et } q \text{ et } r]$

$$I^0(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1}) = 0,220$$

$$I^1(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1}) = 0,351$$

.information apportée par la proposition $[p \text{ et } q \text{ et } r \text{ et } s]$

$$I^0(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1} \cap \hat{C}_{\beta_2}) = 0,386$$

$$I^1(\hat{C}_{\alpha_1} \cap \hat{C}_{\alpha_2}^c \cap \hat{C}_{\beta_1} \cap \hat{C}_{\beta_2}) = 0,492$$

Nous obtenons donc ce que nous avons souhaité ; l'information apportée par la proposition $[p \text{ et } q \text{ et } r \text{ et } s]$ est strictement plus grande que celle de la proposition $[p \text{ et } q \text{ et } r]$ sans que cela s'accompagne nécessairement d'une meilleure localisation dans Ω .

Nous avons introduit une hypothèse très restrictive dans ce paragraphe : les observateurs de O_1 sont supposés indépendants ; on peut évidemment obtenir des résultats plus fins sans cette hypothèse en utilisant la technique indiquée dans les paragraphes II.3, III.2 et VI.3.

VII.1.2 - INFORMATIONS H ET K SUR $O = O_1 \cup O_2$.

Avec les notations définies au début des paragraphes VII.1 et les hypothèses des paragraphes VI.5.3, 4 et 5 nous avons :

$$H(\hat{\Pi}_{O_i}) = K(\hat{\Pi}, O_i) = \sum_{\xi \in O_i} \lambda(\xi) \cdot H_{\xi}(\Pi_{\xi}) \quad i = 1, 2$$

Prenons comme pondération λ des observateurs de O_1 celle qui est définie dans le tableau λ_1 ($\lambda_1(\xi) = 0,1$ pour tout $\xi \in O_1$) ; et définissons les informations H_{ξ}^0 et H_{ξ}^1 à partir des informations I_{ξ}^0 et I_{ξ}^* des paragraphes VI.2.3. et VI.1.4.

Nous attachons l'indice 1 aux informations construites à partir des informations au sens de SHANNON (logarithmes en base 2) et l'indice 0 pour celles qui sont construites à partir des résultats du chapitre IV. On obtient

$$H^0(\hat{\Pi}_{O_1}) = K^0(\hat{\Pi}, O_1) = 1$$

$$H^1(\hat{\Pi}_{O_1}) = K^1(\hat{\Pi}, O_1) = 0,956.$$

Nous avons de même avec λ_2 sur 0_2 :

$$H^0(\hat{\pi}_{0_2}) = K(\hat{\pi}, 0_2) = 1$$

$$H^1(\hat{\pi}_{0_2}) = K(\hat{\pi}, 0_2) = 0,951$$

On sait que, $j = 0, 1$:

$$K^j(\hat{\pi}, 0) = K^j(\hat{\pi}, 0_1) + K^j(\hat{\pi}, 0_2) - K^j(\hat{\pi}, (0_1, 0_2))$$

$$H^j(\hat{\pi}) = H^j(\hat{\pi}_{0_1}) + H^j(\hat{\pi}_{0_2}) - H^j(\hat{\pi}_{0_1}, \hat{\pi}_{0_2})$$

avec

$$K^j(\hat{\pi}, (0_1, 0_2)) = H^j(\hat{\pi}_{0_1}, \hat{\pi}_{0_2}) = \sum_{(\varepsilon_1, \varepsilon_2) \in 0_1 \times 0_2} \lambda(\varepsilon_1, \varepsilon_2) \cdot H^j_{\varepsilon_1 \varepsilon_2}(\hat{\pi}_{\varepsilon_1}, \hat{\pi}_{\varepsilon_2}).$$

En prenant le tableau λ_{12} pour définir la pondération sur $(0_1 \times 0_2, F_1 \otimes F_2)$, il vient :

$$H^0(\hat{\pi}_{0_1}, \hat{\pi}_{0_2}) = K^0(\hat{\pi}, (0_1, 0_2)) = 0,666$$

$$H^1(\hat{\pi}_{0_1}, \hat{\pi}_{0_2}) = K^1(\hat{\pi}, (0_1, 0_2)) = 0,577$$

ce qui nous donne donc

$$H^0(\hat{\pi}) = K^0(\hat{\pi}, 0) = 1,334$$

$$H^1(\hat{\pi}) = K^1(\hat{\pi}, 0) = 1,330$$

VII.1.3 - INFORMATION MUTUELLE $H(\hat{\Pi}_{0_1}; \hat{\Pi}_{0_2})$.

Les tableaux de pondération λ_2 et λ_{12} que nous avons choisis dans le paragraphe précédent ne sont pas quelconques. Il est facile de voir que, dans l'exemple précis que nous étudions, étant donné $\Omega \times O_1$, $\Omega \times O_2$ et λ_1 , les tableaux de pondération λ_2 et λ_{12} sont les meilleurs dans le sens suivant : ils maximisent l'information mutuelle $H^j(\hat{\Pi}_{0_1}; \hat{\Pi}_{0_2})$, $j = 0,1$, entre O_1 et O_2 pour l'observation de $\hat{\Pi}$, ou, ce qui revient au même, ils minimisent l'information conditionnelle apportée sur $\hat{\Pi}$ par un groupe d'observateurs si l'on connaît l'information apportée sur $\hat{\Pi}$ par l'autre groupe d'observateurs.

Nous savons (VI.5.5.) que : $j = 0,1$

$$0 \leq H^j(\hat{\Pi}_{0_1}; \hat{\Pi}_{0_2}) \leq H^j(\hat{\Pi}_{0_i}) \quad i = 1,2$$

L'information mutuelle est nulle si les couples d'observateurs $(\xi_1, \xi_2) \in O_1 \times O_2$ tels que $\lambda_{12}(\xi_1, \xi_2) > 0$ sont indépendants en information. Cette quantité est d'autant plus grande que ces couples d'observateurs sont plus "dépendants" en information ; en particulier, dans notre exemple, nous aurions $H^j(\hat{\Pi}_{0_1}; \hat{\Pi}_{0_2}) = H^j(\hat{\Pi}_{0_2})$, $j = 0,1$, si pour tous les couples

$(\xi_1, \xi_2) \in O_1 \times O_2$ tels que $\lambda_{12}(\xi_1, \xi_2) > 0$ les observateurs étaient identiques pour l'observation de Ω , c'est-à-dire avec les pondérations définies, si :

$$\alpha = \alpha_1 = \alpha_2 ; \quad \beta = \beta_1 = \beta_2 = \beta_3 = \beta_4 ; \quad \gamma = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$$

pour l'observation de Ω .

Nous avons obtenu

$$0 \leq H^0(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2}) = 0,666 \leq H^0(\hat{\pi}_{0_2}) = 1$$

$$0 \leq H^1(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2}) = 0,577 \leq H^1(\hat{\pi}_{0_2}) = 0,951$$

Etant donné la mesure H^0 , une quantité d'information mutuelle égale à 0,666 est-elle l'indice d'une liaison faible, moyenne, forte ?

La solution générale à ce type de question passe d'abord par l'étude du problème indiqué au paragraphe VI.2.6. c ; cependant, pour fixer nos idées, nous avons utilisé un processus de simulation.

Etant donné le tableau $\Omega \times \mathcal{O}_2$ des trois observateurs $\mathcal{O}_2 = \{\alpha, \beta, \gamma\}$, nous simulons le tableau $\Omega \times \mathcal{O}_1$ de la manière suivante: soit $0 \leq p_0 \leq 1$

α_1, α_2 sont des "répliques" de α à " p_0 près"
 $\beta_1, \beta_2, \beta_3, \beta_4$ sont des "répliques" de β à " p_0 près"
 $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ sont des "répliques" de γ à " p_0 près"

Précisément, nous avons mis en place une procédure aléatoire telle que, par exemple :

- pour ω_1 la réponse à l'observateur $\alpha \in \mathcal{O}_2$ vaut 1 ; la réponse simulée pour l'observateur $\alpha_1 \in \mathcal{O}_1$ sera donc égale à 1 avec la probabilité $(1 - p_0)$ et à 0 avec la probabilité p_0 .
- pour ω_{10} la réponse à l'observateur $\alpha \in \mathcal{O}_2$ vaut 0 ; la réponse simulée pour l'observateur $\alpha_1 \in \mathcal{O}_1$ sera donc égale à 0 avec la probabilité $(1 - p_0)$ et à 1 avec la probabilité p_0 .

Les 160 variables de Bernoulli que nous utilisons (pour chaque $\omega \in \Omega$ et chaque $\xi \in \mathcal{O}_1$) sont indépendantes.

Pour chaque valeur de p_0 , nous avons simulé ainsi 50 tableaux $\Omega \times \mathcal{O}_1$ et calculé à chaque fois $H^j(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2})$, $j = 0, 1$, (avec la même mesure de pondération λ_{12}) ; nous avons ensuite déterminé la moyenne $\bar{H}^j(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2})$ et l'écart-type $\sigma(H^j(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2}))$ de $H^j(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2})$, $j = 0, 1$, pour ces 50 tirages.

Les résultats sont indiqués dans le tableau de la page VII.13. Nous avons limité p_0 à l'intervalle $\left[0, \frac{1}{2}\right]$ car il y a évidemment symétrie des résultats obtenus avec $p_0 \in \left[0, \frac{1}{2}\right]$ ou avec $p'_0 = (1 - p_0) \in \left[\frac{1}{2}, 1\right]$ dans notre exemple.

Si l'on admet que le tableau $\Omega \times \mathcal{O}_1$ que nous avons étudié a été réellement défini à partir de $\Omega \times \mathcal{O}_2$ suivant le processus indiqué, on voit que les valeurs :

$$H^0(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2}) = 0,666$$

$$\text{ou } H^1(\hat{\pi}_{0_1} ; \hat{\pi}_{0_2}) = 0,577$$

conduisent à penser que, dans ce cas, p_0 devrait être voisin de 10 %.

P_0	$\bar{H}^0$	$\sigma(H^0)$	$\bar{H}^1$	$\sigma(H^1)$
0	1	0	0.951	0
0.2	0.921	0.038	0.860	0.044
0.4	0.859	0.050	0.792	0.055
0.6	0.804	0.057	0.734	0.063
0.8	0.706	0.060	0.628	0.063
0.10	0.662	0.071	0.584	0.093
0.12	0.600	0.077	0.523	0.079
0.14	0.553	0.068	0.477	0.066
0.16	0.501	0.083	0.429	0.079
0.18	0.452	0.079	0.384	0.078
0.20	0.390	0.079	0.327	0.074
0.25	0.296	0.059	0.244	0.051
0.30	0.214	0.049	0.174	0.044
0.35	0.149	0.043	0.119	0.029
0.40	0.109	0.042	0.086	0.034
0.45	0.077	0.032	0.060	0.025
0.50	0.073	0.027	0.057	0.024

Moyenne et écart-type des informations mutuelles H^0 et H^1
 en fonction de p_0 (§ VII.1.3).



VII.1.4 - INFORMATION SUR UNE PARTITION.

Dans cette partie, nous essayons de montrer clairement sur l'exemple ce que les notions d'information moyenne que nous avons introduites apportent par rapport à l'information classique au sens de SHANNON.

VII.1.4 a) INFORMATION H^* AU SENS DE SHANNON.

Considérons d'abord que les tableaux $\Omega \times \mathcal{O}_1$, $\Omega \times \mathcal{O}_3$, $\Omega \times \mathcal{O}_4$, $\Omega \times \mathcal{O}_5$ de la page VII.2 définissent des partitions de Ω que nous notons Π_1 , Π_3 , Π_4 , Π_5 , c'est-à-dire avec les conventions indiquées au début de ce paragraphe VII.1 :

$$\Pi_i = T_i^{-1}(\hat{\Pi}_{\mathcal{O}_i}) \quad i = 1, 3, 4, 5.$$

Nous avons sur (Ω, S) , en notant $\leq$ la relation de finesse définie en VI.1.1 :

$$\Pi_1 \leq \Pi_3 \leq \begin{cases} \Pi_4 \\ \Pi_5 \end{cases}.$$

Il vient :

$$H^*(\Pi_1) = 3,875$$

$$H^*(\Pi_3) = 2,430$$

$$H^*(\Pi_4) = 1,924$$

$$H^*(\Pi_5) = 1,924.$$

Nous avons évidemment pour $i = 3, 4, 5$

$$H^*(\Pi_1 ; \Pi_i) = H^*(\Pi_i) \quad \text{car} \quad \Pi_1 \leq \Pi_i$$

enfin les partitions Π_4 et Π_5 ayant le même nombre d'éléments et une distribution de probabilités identiques, nous avons obtenu

$$H^*(\Pi_4) = H^*(\Pi_5).$$

Donc pour l'information au sens de SHANNON les partitions Π_4 et Π_5 sont INDISCERNABLES aussi bien pour l'information en soi que pour l'information mutuelle par rapport à Π_1 .

Par définition des Π_i , $i = 1, 3, 4, 5$ et pour les raisons déjà indiquées au paragraphe VII.1.1. a), on obtient les mêmes résultats en se plaçant dans les espaces $(\hat{\Omega}_i, \hat{S}_i)$

$$H^*(\hat{\Pi}_{0_i}) = H^*(\Pi_i) \quad i = 1, 3, 4, 5$$

et dans les espaces $(\hat{\Omega}_1 \times \hat{\Omega}_i, \hat{S}_1 \otimes \hat{S}_i)$ $i = 3, 4, 5$

$$H^*(\hat{\Pi}_{0_1}; \hat{\Pi}_{0_i}) = H^*(\Pi_1; \Pi_i) \quad i = 3, 4, 5.$$

Il n'y a aucune différence entre l'information au sens de SHANNON sur (Ω, S) ou sur les espaces $(\hat{\Omega}_i, \hat{S}_i)$, $i = 1, 3, 4, 5$.

VII.1.4 b) INFORMATION H^0 et H^1 .

Reprenons les calculs explicités au paragraphe VII.1.2. en remplaçant 0_2 respectivement par $0_3, 0_4, 0_5$; il vient :

$$H^1(\hat{\Pi}_{0_1}) = 0,957$$

$$H^0(\hat{\Pi}_{0_1}) = 1$$

$$H^1(\hat{\Pi}_{0_3}) = 2,430$$

$$H^0(\hat{\Pi}_{0_3}) = 5$$

$$H^1(\hat{\Pi}_{0_4}) = 1,924$$

$$H^0(\hat{\Pi}_{0_4}) = 3$$

$$H^1(\hat{\Pi}_{0_5}) = 1,924$$

$$H^0(\hat{\Pi}_{0_5}) = 3$$

et aussi

$$H^1(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_3}) = 0,721$$

$$H^0(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_3}) = 0,765$$

$$H^1(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_4}) = 0,545$$

$$H^0(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_4}) = 0,599$$

$$H^1(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_5}) = 0,410$$

$$H^0(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_5}) = 0,443$$

Nous n'avons plus les relations du type

$$H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_i}) = H^j(\hat{\Pi}_{0_i})$$

$$j = 0,1$$

$$i = 3,4,5$$

que nous avions dans le paragraphe précédent car les relations de finesse

$$\hat{\Pi}_{0_1} \leq \hat{\Pi}_{0_i}$$

$$i = 3,4,5$$

n'ont pas de sens dans $\hat{\Omega}_1 \times \hat{\Omega}_i$ $i = 3,4,5$.

Cependant nous avons le résultat original suivant ; on peut considérer $\hat{\Pi}_{0_3}, \hat{\Pi}_{0_4}, \hat{\Pi}_{0_5}$ comme étant trois partitions du même espace (Ω, S) ; nous avons alors

$$\hat{\Pi}_0 \leq \hat{\Pi}_{0_3} \leq \hat{\Pi}_{0_4}$$

$$\hat{\Pi}_0 \leq \hat{\Pi}_{0_3} \leq \hat{\Pi}_{0_5}$$

en notant $\hat{\Pi}_0$ la partition la plus fine de Ω .

Nous avons montré (VI.5.5) que $\hat{\Pi}_{0_1}$ étant donné, les applications

$H^j(\hat{\Pi}_{0_1} ; \cdot)$ définissent des informations sur les partitions de (Ω, S) ; en effet nous avons pour $j = 0, 1$

$$0 \leq \left\{ \begin{array}{l} H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_4}) \\ H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_5}) \end{array} \right\} \leq H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_3}) \leq H^j(\hat{\Pi}_{0_1}) = H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_1}).$$

Notons de plus que

- $H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}')$ est maximum et égal à $H^j(\hat{\Pi}_{0_1})$ si et seulement si $T_1^{-1}(\hat{\Pi}_{0_1}) \geq \hat{\Pi}'$.
- $H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}')$ est nulle si et seulement si toutes les partitions Π_ξ , $\xi \in O_1$, sont indépendantes en information de $\hat{\Pi}'$.

Le résultat fondamental et original des informations que nous avons introduites au chapitre VI par rapport à l'information classique de SHANNON est celui-ci :

à $\hat{\Pi}_{0_4}$ et $\hat{\Pi}_{0_5}$ sont associées des partitions de (Ω, S) qui comportent le même nombre d'éléments et définissent une même répartition de probabilités sur ces éléments ; ils sont donc indiscernables au sens de l'information de SHANNON comme nous l'avons signalé (VII.1.4. a). Pourtant nous leur associons une quantité d'information mutuelle différente. Dans notre exemple si :

$$H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_4}) > H^j(\hat{\Pi}_{0_1} ; \hat{\Pi}_{0_5}) \quad j = 0, 1$$

c'est parce que la partition $\hat{\Pi}_{0_4}$ "résume mieux" que $\hat{\Pi}_{0_5}$ la partition $\hat{\Pi}_{0_1}$.

Intuitivement, *vu les données*, il est clair que si l'on cherche à définir les réponses d'un ω quelconque dans $\hat{\Omega}_1$ dire :

" ω appartient à l'ensemble des 1 de $\hat{\Pi}_{0_4}$ "

c'est apporter une information sur les réponses de ω plus importante que dire

" ω appartient à l'ensemble des 1 de $\hat{\Pi}_{0_5}$ ".

En effet, dans $\Omega \times \mathcal{O}_1$ affirmer

" $\omega \in \{\omega_1, \omega_2, \omega_3, \omega_{15}, \omega_{16}\}$ "

c'est apporter moins d'information que d'affirmer

" $\omega \in \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ "

car dans ce dernier cas les "profils" de réponses des $\{\omega_i | i = 1, \dots, 5\}$ se ressemblent beaucoup plus que ceux des $\{\omega_i | i = 1, 2, 3, 15, 16\}$;

c'est ce que traduit l'inégalité précédente entre les quantités d'information mutuelle.

Nous allons axer la présentation des perspectives d'applications du paragraphe suivant sur cette propriété.

VII.2 - HIERARCHIE - CLASSIFICATION - QUESTIONNAIRE.

Nous présentons dans ce paragraphe quelques directions de travail en vue d'applications, en essayant de mettre en relief l'apport des notions introduites précédemment pour l'étude de certains problèmes de l'analyse des données [3], [22], [33], ou de la théorie des questionnaires [26], [36].

Les ordinateurs sont pour les utilisateurs de puissants moyens de calcul. Ils leur donnent la possibilité de s'intéresser à de vastes tableaux de données qualitatives ou quantitatives. Le but des différentes méthodes

d'analyse de ces données consiste essentiellement à extraire le maximum d'information de ces tableaux et à présenter cette information résumée sous une forme facile à appréhender. Pour comparer différentes méthodes ou différents résultats d'une même méthode, il faut, semble-t-il, pouvoir maîtriser deux critères. D'une part, un critère qui quantifie la valeur de l'information résumée ; non pas la valeur intrinsèque, *mais la valeur de l'information que le résumé apporte sur l'information intégrale des données de départ* ; quand les données et le résumé statistiques peuvent être présentés sous la forme d'une partition ou d'un ensemble de partitions, il est clair que la notion d'information mutuelle entre deux groupes d'observateurs, que nous avons étudiée au chapitre précédent, apporte une première solution à ce problème. D'autre part, il est utile de maîtriser, dans certains cas, la notion de *coût du résumé*, de manière à pouvoir, dans la pratique, faire un choix en toute connaissance de cause entre deux systèmes en comparant à la fois leurs informations et leurs coûts respectifs.

On trouve déjà quelques éléments de réflexion sur ces sujets dans l'exemple détaillé du paragraphe précédent auquel nous nous référerons constamment.

VII.2.1 - UNE METHODE DE CLASSIFICATION HIERARCHIQUE DESCENDANTE.

Dans l'exemple que nous avons traité au paragraphe VII,1 il est clair que les tableaux $\Omega \times O_2$ et $\Omega \times O_3$, que nous avons définis a priori, sont intuitivement de "bons résumés" de $\Omega \times O_1$. *Comment dans la pratique, étant donné le tableau $\Omega \times O_1$, déterminer les observateurs de O_2 et la partition $\hat{\Pi}_{O_2}$ qui résument aussi bien l'information ?*

Nous proposons d'utiliser l'algorithme suivant : étant donné le tableau $\Omega \times O_1$, la mesure de probabilité λ_1 sur (O_1, F_1) et l'ensemble des mesures d'information $\{I_{\xi_1 \xi_2} \mid (\xi_1, \xi_2) \in O_1 \times O_1\}$ pour tous les couples d'observateurs de O_1 , déterminons successivement :

A la 1ère étape : l'observateur $\xi_1 \in O_1$ qui apporte le plus d'information sur $\Omega \times O_1$ au sens du paragraphe VII.1.4. b).

A la n^{ème} étape : le sous-ensemble O_2^n de O_1 formé des $n-1$ observateurs O_2^{n-1} de l'étape précédente et de l'observateur $\xi_n \in O_1 - O_2^{n-1}$ tel que la partition $\hat{\Pi}_{O_2^n}$ de Ω apporte le plus d'information sur $\Omega \times O_1$ (au sens du paragraphe VII.1.4. b).

On trouve à la page VII.21 l'application de l'algorithme que nous venons de décrire succinctement aux données du tableau $\Omega \times O_1$ de l'exemple précédent pour la mesure d'information H^1 (VII.1.4 b). Les résultats sont présentés sous la forme d'une arborescence représentative de la hiérarchie de parties ainsi définie. Nous avons indiqué, à chaque niveau de l'arbre, l'observateur sélectionné et l'information mutuelle entre la partition $\hat{\Pi}$ de Ω ainsi définie et le tableau $\Omega \times O_1$ de départ (c'est-à-dire $H^1(\hat{\Pi}_{O_1}; \hat{\Pi})$ au sens du paragraphe VII.1.4. b).

Le premier observateur sélectionné est β_2 et l'on a

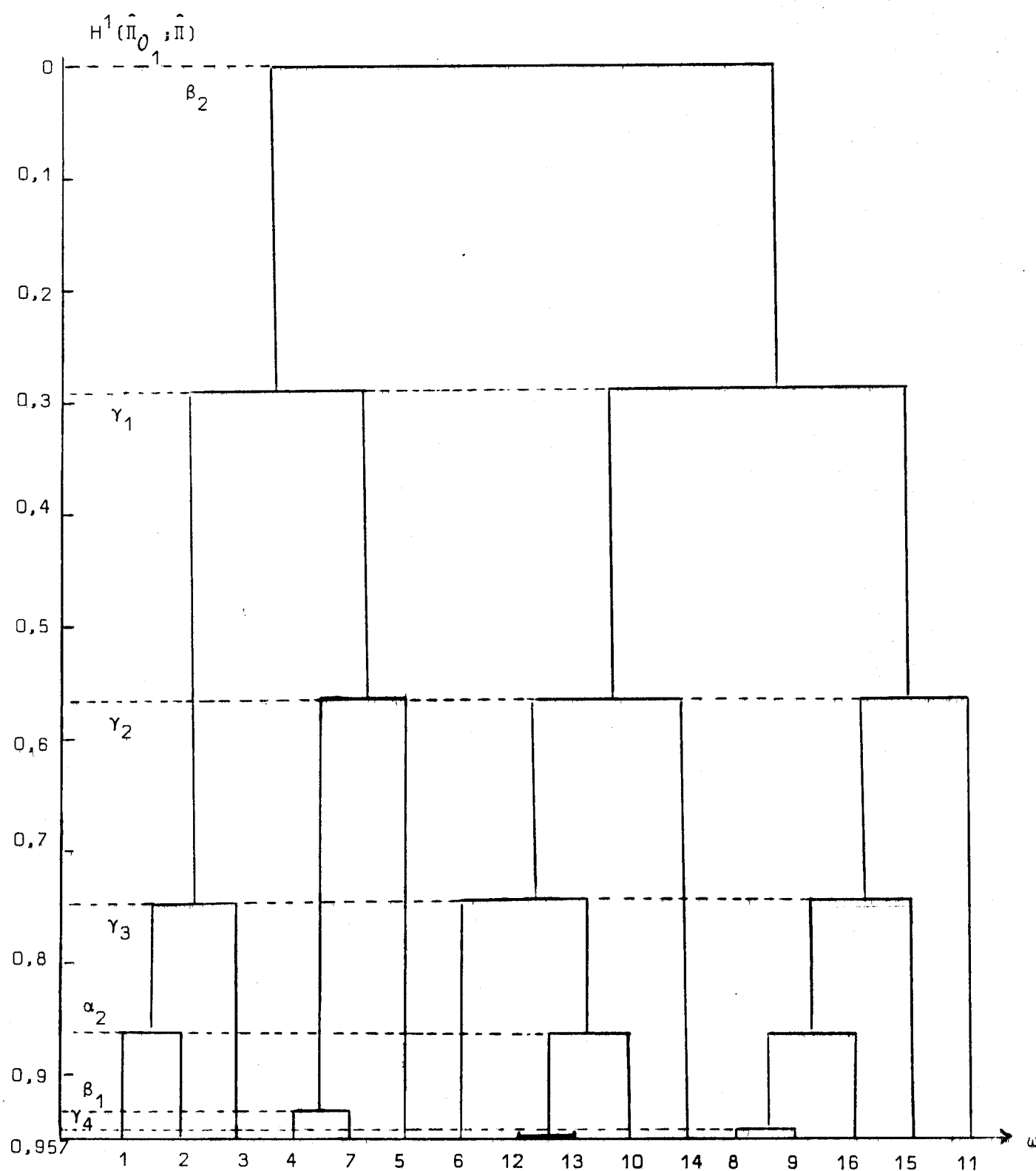
$$H^1(\hat{\Pi}_{O_1}; \hat{\Pi}) = 0,28 \quad \hat{\Pi} = \hat{\Pi}_{\beta_2}.$$

Le deuxième observateur sélectionné est γ_1 et l'on a

$$H^1(\hat{\Pi}_{O_1}; \hat{\Pi}) = 0,57 \quad \hat{\Pi} = \hat{\Pi}_{\beta_2} \cap \hat{\Pi}_{\gamma_1}$$

etc...

Notons que la hiérarchie de parties ainsi obtenue n'est probablement pas optimum ; par exemple ω_7 et ω_{11} semblent mal placés dans la classification. On voit facilement que cela tient à l'existence de 0 ou de 1 parasites dans leurs profils. On peut modifier l'algorithme de manière à



CLASSIFICATION HIERARCHIQUE DE L'ENSEMBLE Ω (§ VII.2.1.)



déterminer à chaque étape un observateur (par exemple β_2 à la première étape) et ceux qui lui sont fortement corrélés en information ($\beta_1, \beta_3, \beta_4$) et déduire, de ce groupe d'observateurs, la partition à réaliser.

Il conviendrait évidemment de comparer le résultat de cet algorithme à ceux d'autres méthodes, et surtout sur des ensembles beaucoup plus vastes de données correctes. Notons cependant que nous avons sélectionné un observateur de type β et un observateur de type γ , ce qui représente bien les deux composantes les plus importantes du tableau $\Omega \times \mathcal{O}_1$. Remarquons aussi que cet algorithme fournit à la fois une classification et une "clé" pour cette classification.

En analyse des données, quand les résumés statistiques peuvent être présentés sous la forme d'une partition ou d'un ensemble de partitions, la notion d'information mutuelle apparaît donc comme un moyen de comparer les résultats de différentes méthodes de classification ou les différents résultats d'une même méthode, car elle permet de mesurer, dans chaque cas, la valeur de l'information des résumés ainsi obtenus. L'étude de l'information mutuelle conduit aussi à proposer des algorithmes originaux de classification.

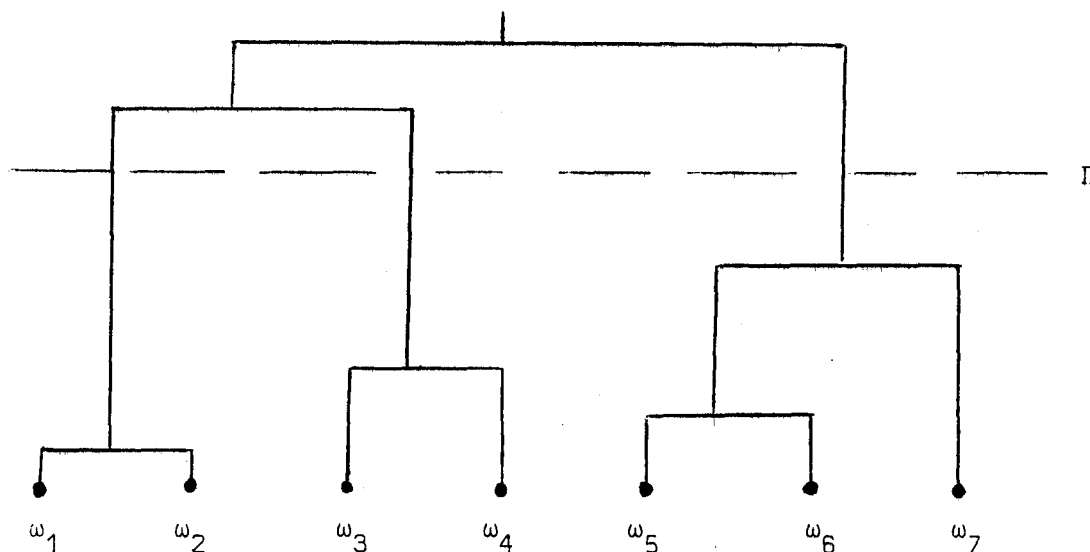
Nous commençons à entrevoir par ce biais des solutions nouvelles au problème très général qui nous a été souvent posé de la manière suivante [34], [35] : *après avoir étudié un échantillon représentatif d'une population de bactéries au moyen d'une centaine de tests, le biologiste nous demande de l'aider, d'une part à déterminer une classification de ces bactéries, d'autre part à sélectionner un sous-ensemble d'une vingtaine de tests qui permettent facilement, étant donné une nouvelle bactérie, de déterminer sa place dans la classification ainsi définie. Les vingt tests sélectionnés doivent être ceux qui apportent "le plus d'information", compte tenu, le cas échéant, du coût de chaque test. L'algorithme précédent est une méthode, parmi d'autres, pour sélectionner ces*

vingt tests et définir la classification des bactéries : nous allons proposer dans le paragraphe VII.2.3. une autre méthode de "diagnostic", c'est-à-dire une méthode pour déterminer la place d'un nouvel élément dans la classification.

On rencontre ce type de problème chaque fois que l'on désire étudier de manière exhaustive un petit échantillon afin de déterminer quelques indicateurs pertinents et peu coûteux que l'on utilisera pour l'étude d'un grand échantillon ou de la population totale.

VII.2.2 - INFORMATION ET HIERARCHIE DE PARTIES.

Supposons que le résultat d'un algorithme de classification hiérarchique ou d'un questionnaire [26] soit une arborescence représentative d'une hiérarchie de parties, c'est-à-dire une succession de partitions de plus en plus fines d'un ensemble Ω (qui sera par exemple un ensemble d'individus statistiques en analyse des données ou un ensemble d'éventualités dans un questionnaire).



Des coupes convenablement définies dans cet arbre permettent de réaliser des partitions de l'ensemble $\Omega = \{\omega_1, \dots, \omega_7\}$. Par exemple :

$\Pi = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5, \omega_6, \omega_7\}\}$. Nous pouvons attacher à chaque partition Π de Ω deux quantités d'information quand la base de départ est un tableau de

données $\Omega \times \emptyset$ de la forme du tableau $\Omega \times \emptyset_1$ de l'exemple du paragraphe VII.1 :

- d'une part, l'information mutuelle entre la partition Π et le tableau (au sens où nous la calculons dans le paragraphe VII.1.4. b) que nous noterons $H(\Pi; \emptyset)$ c'est-à-dire l'information apportée par Π sur les données de départ de $\Omega \times \emptyset$.
- d'autre part, une information du type SHANNON sur la partition Π que nous noterons $H^*(\Pi)$. Cette quantité peut s'interpréter comme la complexité du modèle Π (plus la partition Π est fine, plus l'information qu'elle apporte est difficile à appréhender pour l'utilisateur).

Nous savons (VII.1.4.b) que pour deux partitions Π_1 et Π_2 de Ω comportant le même nombre d'éléments et ayant la même répartition de probabilités, nous avons

$$H^*(\Pi_1) = H^*(\Pi_2)$$

mais que l'on peut aussi avoir : $H(\Pi_1; \emptyset) \neq H(\Pi_2; \emptyset)$

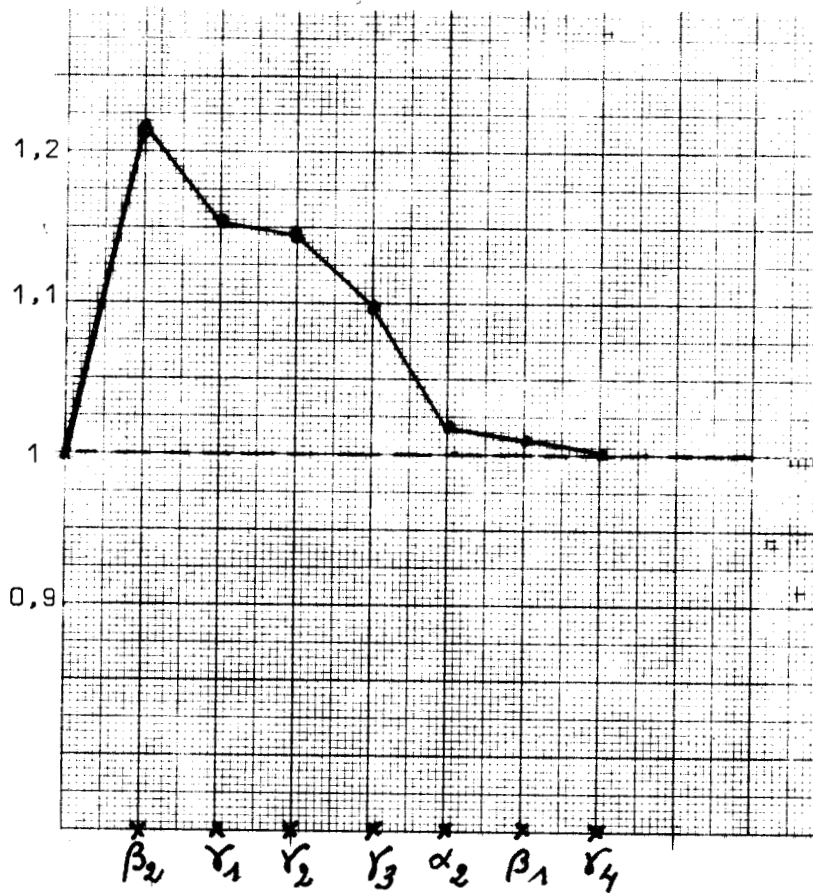
c'est-à-dire que pour *la même complexité des deux modèles (au sens de l'information de SHANNON) nous pouvons déterminer le modèle qui apporte le plus d'information sur les données de départ.*

En comparant, pour chaque coupe de l'arbre, ces deux quantités (éventuellement convenablement normées) on peut définir des positions d'équilibre local, intéressantes à mettre en relief. Par exemple, pour une partition Π_0 donnée, montrer que :

- les partitions Π plus fines que Π_0 (et qui la précèdent immédiatement) apportent un gain d'information (au sens $H(\Pi; \emptyset)$) faible vis à vis de la complexité croissante du modèle (au sens $H^*(\Pi)$).
- les partitions Π plus grossières que Π_0 (et qui lui succèdent directement)

sont moins complexes (au sens $H^*(\Pi)$) mais font perdre une partie trop importante de l'information (au sens $H(\Pi; \emptyset)$) pour être retenues.

Dans l'exemple du paragraphe VII.2.1. précédent nous avons calculé pour chaque coupe de l'arbre d'une part $A(\hat{\Pi}) = H^1(\hat{\Pi}_{0_1}; \hat{\Pi}) / H^1(\hat{\Pi}_{0_1})$ d'autre part, $B(\hat{\Pi}) = H^*(\hat{\Pi}) / H^*(\hat{\Pi}_{0_1})$, et nous avons tracé le graphique ci-dessous des variations du rapport $\frac{A(\Pi)}{B(\Pi)}$ en fonction du nombre d'observateurs sélectionnés.



VARIATION DE $\frac{A(\hat{\Pi})}{B(\hat{\Pi})}$ EN FONCTION DU NOMBRE D'OBSERVATEURS SÉLECTIONNÉS.

(§ VII.2.2).

L'étude de ce graphique nous indique clairement que, d'une part, l'essentiel de la classification réside dans le partage de Ω en deux ensembles par l'observateur β_2 , d'autre part que la partition engendrée par $(\beta_2, \gamma_1, \gamma_2)$ et éventuellement γ_3 est un bon résumé du tableau $\Omega \times O_1$, car affiner cette partition apporte un gain d'information faible vis à vis de la complexité croissante du modèle. Ici encore, seules une expérimentation systématique de ce type de méthode sur des exemples concrets et la comparaison avec les méthodes existantes peuvent mettre en évidence l'intérêt de ce genre d'approche.

VII.2.3. - QUESTIONNAIRES.

On peut, le cas échéant, introduire une fonction coût sur l'ensemble F_1 des groupes d'observateurs de O_1 et, à chaque étape de la méthode proposée dans le paragraphe VII.2.1, ne pas choisir obligatoirement l'observateur qui apporte le plus d'information, mais plutôt celui dont le rapport information-coût est le meilleur. La méthode de classification devient alors un moyen de sélectionner un sous-ensemble O_2 d'observateurs de O_1 qui optimise ce critère.

Une autre modification possible consiste, à chaque étape, à considérer la partition de Ω qui est engendrée par les observateurs sélectionnés, et à choisir pour chaque ensemble de cette partition l'observateur qui apporte le plus d'information. C'est-à-dire à considérer chaque élément de la partition comme un ensemble Ω à étudier par le sous-ensemble des observateurs de O_1 qui n'ont pas été utilisés pour le définir.

C'est l'algorithme que nous avons employé pour obtenir l'arbre de la page VII.28 à partir de l'exemple $\Omega \times O_1$ du paragraphe VII.1.

Nous choisissons d'abord l'observateur de O_1 qui apporte le plus d'information sur $\Omega \times O_1$ (au sens du paragraphe VII.1.4.b), c'est-à-dire β_1 qui divise Ω en deux ensembles $A_{\beta_1} = \{\omega_1, \dots, \omega_5\}$ et $\bar{A}_{\beta_1} = \{\omega_6, \dots, \omega_{16}\}$ puis dans chaque groupe A_{β_1} et $\bar{A}_{\beta_1}$ nous recommençons le processus.

Pour diviser A_{β_1} c'est l'observateur γ_1 qui apporte le plus d'information etc...

Nous avons utilisé, pour construire l'arbre de la page VII.28, l'information mutuelle H^0 au sens du paragraphe VII.1.4.b.

Nous avons indiqué pour chaque noeud de l'arbre l'observateur qui sert à partitionner l'ensemble étudié et les valeurs 0 et 1 qui sont associées à chaque élément de la partition ainsi formée. Les ordonnées de chaque noeud, et celles de chaque sommet, indiquent l'information nécessaire pour "arriver" au point en question.

Par exemple l'ordonnée de ω_3 ($= 0,48$) indique que l'information mutuelle $H^0(\hat{\Pi}_{O_1} ; \hat{\Pi}_{\omega_3})$ entre $\hat{\Pi}_{O_1}$ et la partition $\hat{\Pi}_{\omega_3} = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4, \omega_5\}, \{\omega_6, \dots, \omega_{16}\}\}$ nécessaire pour sélectionner $\hat{\Pi}_3$ "coûte" 0,48 en information mutuelle.

De la même manière la partition $\hat{\Pi}_3$ en trois éléments

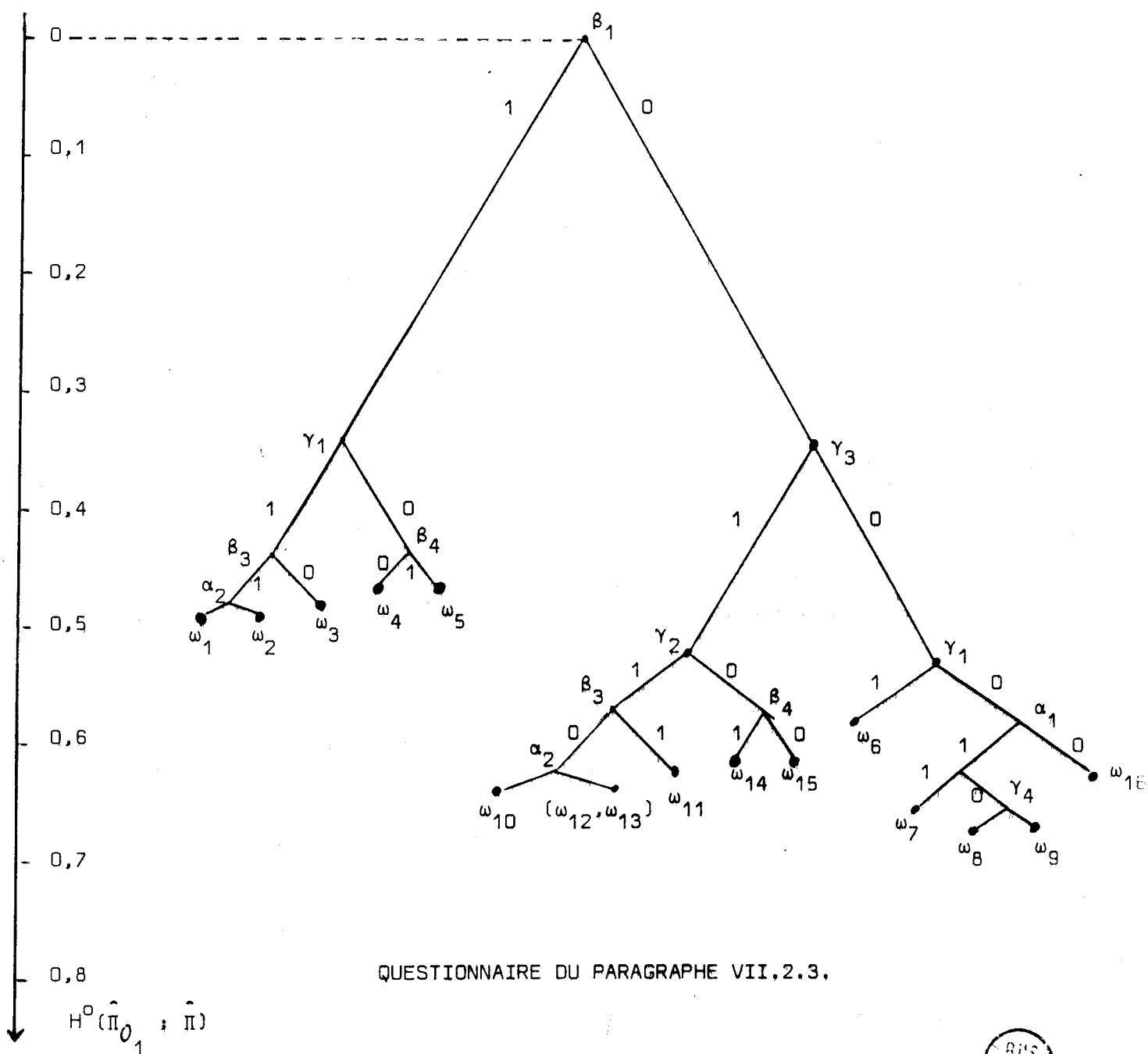
$$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$$

$$\{\omega_{10}, \omega_{11}, \omega_{12}, \omega_{13}, \omega_{14}, \omega_{15}\}$$

$$\{\omega_6, \omega_7, \omega_8, \omega_9, \omega_{16}\}$$

engendrée par l'observateur β_1 et l'observateur γ_3 dans la branche 0 de β_1 , est telle que

$$H^0(\hat{\Pi}_{O_1} ; \hat{\Pi}_3) = 0,52$$



L'utilisation concrète de ce type d'arbre est immédiat : étant donné un élément $\omega \in \Omega$ quelconque, quel critère, c'est-à-dire quel observateur faut-il lui appliquer pour obtenir le maximum d'information sur ω ? A chaque étape, nous ne cherchons pas seulement la meilleure localisation de ω , mais plutôt à obtenir le maximum de renseignements sur ω , c'est-à-dire sur son profil dans $(\Omega \times \mathcal{O}_1)$. En fait, nous avons réalisé un questionnaire du type de ceux qui sont développés par C.F. PICARD [26] et ces collaborateurs comparables aux questionnaires avec coût lié à la base, puisque nous ne nous permettons que les questions, c'est-à-dire les partitions, définies par les observateurs de $\mathcal{O}_1$. On peut aussi facilement imaginer une modification de cet algorithme pour l'utilisation dans un domaine proche de celui des pseudo-questionnaires [36] où les sommets de l'arbre ne sont pas en fait des éléments de $\omega \in \Omega$ mais des répartitions de probabilités sur Ω ou sur des parties de Ω .

Remarquons enfin que si l'on utilise le même algorithme avec l'information classique au sens de SHANNON, on sélectionne, non pas un observateur qui "résume bien" d'autres observateurs, mais celui qui découpe l'ensemble Ω en deux parties aussi égales que possible, par exemple α_2 , γ_1 ou γ_3 pour la première étape de cet algorithme.

VII.2.4 - ANALYSE FACTORIELLE DES CORRESPONDANCES.

Les résultats du chapitre IV permettent de définir des distances entre événements. En particulier, une méthode d'analyse de données développée par J.P. BENZECRI *l'analyse factorielle des correspondances* [3] qui est définie à partir de la distance du CHI-2 entre distributions conditionnelles de probabilités, peut-être aussi définie à partir de distances entre distributions conditionnelles d'informations, comme nous l'avons montré dans [23].

On sait que cette méthode d'analyse factorielle, rigoureusement définie dans le cadre d'une variable statistique à deux dimensions, s'applique

aussi avec succès à des tableaux de correspondances dont les données sont bien loin de vérifier ce modèle. Il semble, en liaison avec les problèmes soulevés au paragraphe V.2, que l'on puisse imaginer une nouvelle approche théorique de cette méthode à partir de la théorie de l'information et des résultats précédents.

VII.2.5. - CONCLUSION DU CHAPITRE VII.

Nous n'approfondirons pas d'avantage dans ce mémoire les applications possibles de la théorie généralisée de l'information que nous venons d'évoquer rapidement. Nous pensons cependant avoir suffisamment développé nos idées à partir de l'exemple VII.1 pour montrer que la théorie généralisée de l'information ouvre de nouvelles perspectives d'applications dans les domaines de l'analyse des données et des questionnaires.

Nous croyons que ces applications méritent dès à présent de plus larges développements, mais qu'elles soulèveront aussi de nouvelles difficultés théoriques, en information généralisée, liées aux problèmes que nous avons analysés au chapitre V, et qu'il sera indispensable d'approfondir.

C O N C L U S I O N .

Pour conclure ce mémoire, replaçons notre travail dans un contexte plus général ; pour souligner la distinction qu'il convient de faire entre l'information apportée par *la localisation d'un événement* et l'information apportée par *une interprétation sémantique de cet événement*, J. KAMPE DE FERIET propose l'exemple suivant [19] :

"Si, en visite chez un collègue, je lui demande "La Théorie Analytique de la Chaleur" de FOURIER et qu'il dit successivement :

- p : j'ai ce livre dans ma bibliothèque
- q : il est relié en rouge
- r : c'est un in-quarto
- s : il est sur le 3^{ème} rayon à droite ;

la localisation se précise à chaque nouvelle proposition et mon Information augmente ; elle atteint son maximum au moment où je prends en main ce volume : (*l'objet recherché*) *est localisé* ; mais il est clair que cette Information n'a aucun rapport avec celle qui est *contenue dans le livre* et que me fournirait sa lecture".

La situation décrite est exemplaire. Imaginons en effet, qu'un troisième mathématicien assiste à cet entretien. Ce dialogue lui a fourni deux sortes d'informations ; d'abord une information "sémantique" ; les données du Titre et de l'Auteur du livre désiré lui permettent de définir avec précision le contenu de l'objet recherché ; ensuite une information "localisation" : les réponses du collègue lui permettent de localiser concrètement l'objet-livre dans l'ensemble des objets-livres de la bibliothèque.

Plaçons-nous dans une situation différente : sans idée préconçue sur les titres des ouvrages qui l'intéressent, le visiteur précise successivement qu'il désire consulter les livres :

- p : de Mathématiques
- q : traitant de Statistiques
- r : plus particulièrement des Tests Séquentiels
- s : appliqués au Contrôle de Qualité

Les informations qu'il apporte à son interlocuteur sont des précisions sur le contenu des livres qui l'intéressent. Ces renseignements permettent de localiser de mieux en mieux dans la bibliothèque les livres à lui conseiller ; mais à notre sens on ne peut réduire l'information sémantique qu'il apporte à son interlocuteur en disant :

"Je désire consulter les livres de Mathématiques traitant de Statistiques"

à l'information-localisation associée au sous-ensemble des objets-livres traitant de ce sujet ; même si la bibliothèque est supposée exhaustive ou idéale.

Pourtant, bien des applications possibles de la théorie de l'information se présentent sous cette forme : *ce sont des renseignements sur le contenu sémantique de l'événement qui provoquent, le cas échéant, une localisation de plus en plus précise de cet événement. La notion d'information apportée par un groupe d'observateurs sur un événement permet, dans certains cas simples, d'associer à une série de renseignements une quantité d'information qui est fonction à la fois de la localisation déduite de ces renseignements et de l'intérêt intrinsèque que l'on associe à chaque renseignement.*

Notre modèle ne permet pas de résoudre des problèmes aussi riches et complexes que ceux soulevés par cet exemple. Mais, dans le cas particulier où l'on étudie un ensemble fini d'événements élémentaires décrits par un ensemble fini de caractères, la définition de l'information telle que nous la proposons *dépasse la simple localisation* et permet une quantification de l'information sémantique apportée par une proposition simple du type : *"Les éléments de tel ensemble possèdent tel groupe de propriétés"*.



B I B L I O G R A P H I E

- [1] J. ACZEL LECTURES ON FUNCTIONAL EQUATIONS AND THEIR APPLICATIONS,
Academic Press, New-York-London (1966).
- [2] P. BENVENUTI SULLE MISURE DI INFORMAZIONE COMPOSITIVA CON TRACCIA COMPOSITIVA UNIVERSALE,
Rend. di Mat., VI - 2, pp. 481 - 506 (1969).
- [3] J.P. BENZECRI L'ANALYSE DES DONNEES
et Collaborateurs Tome I : LA TAXINOMIE
Tome II : L'ANALYSE DES CORRESPONDANCES
DUNOD, Paris (1973).
- [4] G. BIRKHOFF LATTICE THEORY,
American Mathematical society, Colloquim Publications, XXV, New-York.(1948).
- [5] R. CARNAP AN OUT-LINE OF A THEORY OF SEMANTIC INFORMATION,
Y. BAR-HILLEL Massachusetts institute of Technology,
Research laboratory of electronics,
Technical Report N° 247 (1952).
- [6] G. COMYN, J. LOSFELD INFORMATION GENERALISEE SUR UNE QUASI-PARTITION ET INFORMATION MOYENNE,
C.R. Acad. Sc. Paris, 276, série A, pp. 1373-1376 (1973).
- [7] G. COMYN, J. LOSFELD APPLICATION DE L'INFORMATION GENERALISEE A L'ANALYSE DES DONNEES STATISTIQUES,
C.R. Acad. Sc. Paris, 276, série A, pp. 1075-1078 (1973).
- [8] B. FORTE INFORMATION AND PROBABILITY : COLLECTORS AND COMPOSITIVITY,
Symposia Mathematica, IX, pp. 121-129 (1972).
- [9] B. FORTE GENERALIZED MEASURES OF INFORMATION AND UNCERTAINTY,
 (Proc. Meeting on Information measures)
KITCHENER-WATERLOO, Ontario, Canada, (Avril 1970).
- [10] B. FORTE SUR UNE CLASSE D'ENTROPIES UNIVERSELLES,
P. BENVENUTI C.R. Acad. Sc. Paris, 268, série A, pp. 1628-1631 (1969).
- [11] B. FORTE, N. PINTACUDA INFORMATION FOURNIE PAR UNE EXPERIENCE
C.R. Acad. Sc. Paris, 266, série A, pp. 242-245 (1968).

- [12] J. KAMPE DE FERIET
B. FORTE INFORMATION ET PROBABILITE,
C.R. Acad. Sc. Paris, 265, série A, pp. 110-114,
pp. 142-146, pp. 350-353 (1967).
- [13] J. KAMPE DE FERIET MESURE DE L'INFORMATION FOURNIE PAR UN EVENEMENT,
Colloques Internationaux, C.N.R.S., 186,
CLERMONT-FERRAND (1969), C.N.R.S. PARIS, pp.191-221
(1970).
- [14] J. KAMPE DE FERIET MESURE DE L'INFORMATION PAR UN ENSEMBLE D'OBSER-
VATEURS,
C.R. Acad. Sc. Paris, 269, série A, pp. 1081-1085
(1969).
- [15] J. KAMPE DE FERIET MESURE DE L'INFORMATION PAR UN ENSEMBLE D'OBSER-
VATEURS,
C.R. Acad. Sc. Paris, 271, série A, pp. 1017-1021
(1970).
- [16] J. KAMPE DE FERIET MESURE DE L'INFORMATION FOURNIE PAR UN EVENEMENT
Séminaire sur les questionnaires, 24 Nov.-8Déc.1971
Structures de l'Information, Institut Henri Poincaré,
PARIS (1972).
- [17] J. KAMPE DE FERIET MESURE DE L'INFORMATION PAR UN ENSEMBLE D'OBSER-
VATEURS INDEPENDANTS,
Transactions of the Sixth Prague Conference on
Information Theory, Statistical decision functions,
Random processes, PRAGUE (1971) Czechoslovak Acad.
of Sciences, PRAGUE, pp. 315-329 (1973).
- [18] J. KAMPE DE FERIET MEASURE OF INFORMATION BY A SET OF INDEPENDENT
OBSERVERS,
Aequationes Mathematicae, 8, pp. 159-161 (1972).
- [19] J. KAMPE DE FERIET LA THEORIE GENERALISEE DE L'INFORMATION ET LA
MESURE SUBJECTIVE DE L'INFORMATION,
Rencontres en Théorie de l'Information,
MARSEILLE-LUMINY, Juin 1973 (à paraître).
- [20] D.A. KAPPOS STRUKTURTHEORIE DER WAHRSCHEINLICHKEITS-FELDER
UND RAÜME,
SPRINGER, BERLIN (1960).
- [21] C. LANGRAND SUR LA MESURABILITE ET L'ESTIMATION.Thèse,LILLE (1968).
- [22] L. LEBART, J.P. FENELON STATISTIQUE ET INFORMATIQUE APPLIQUEES,
DUNOD, PARIS (1971).
- [23] J. LOSFELD INFORMATION MOYENNE DANS UNE EPREUVE STATISTIQUE,
C.R. Acad. Sc. Paris, 275, Série A, pp. 509-512
(1972).
- [24] J. LOSFELD INFORMATION GENERALISEE ET RELATION D'ORDRE,
Rencontres en Théorie de l'Information
MARSEILLE-LUMINY, Juin 1973 (à paraître).

- [25] J. NEVEU BASES MATHÉMATIQUES DU CALCUL DES PROBABILITÉS,
MASSON, PARIS (1964).
- [26] C.F. PICARD GRAPHS ET QUESTIONNAIRES,
TOME I : GRAPHS
TOME II : QUESTIONNAIRES
GAUTHIER-VILLARS, PARIS, (1972).
- [27] E. ROUBINE INTRODUCTION A LA THÉORIE DE LA COMMUNICATION,
THÉORIE DE L'INFORMATION,
Tome III, Masson et Cie, PARIS (1970).
- [28] J. SALLANTIN SYSTÈME DE PROPOSITIONS ET INFORMATIONS,
C.R. Acad. Sc. Paris, 274, Série A, pp. 986-988 (1972).
INFORMATIONS PURES SUR UN SYSTÈME DE PROPOSITIONS,
C.R. Acad. Sc. Paris, 275, Série A, pp. 65-68
(1972).
- [29] J. SALLANTIN INFORMATIONS, SYSTÈMES DE PROPOSITIONS ET LOGIQUE
DE LA MÉCANIQUE QUANTIQUE
Thèse 3^e cycle PARIS 1972.
- [30] B. SCHWEITZER, MESURES ALÉATOIRES DE L'INFORMATION
A. SKLAR C.R. Acad. Sc., PARIS, 269, série A, pp. 721-723
(1969).
- [31] B. SCHWEITZER, MESURE ALÉATOIRE DE L'INFORMATION ET MESURE DE
A. SKLAR L'INFORMATION PAR UN ENSEMBLE D'OBSERVATEURS,
C.R. Acad. Sc., PARIS, 272, série A, pp. 149-153
(1971).
- [32] C. SHANNON, W. WEAVER A MATHEMATICAL THEORY OF COMMUNICATION,
Un. of Illinois Press, URBANA (1949).
- [33] R. SOKAL, P. SNEATH PRINCIPLES OF NUMERICAL TAXONOMY,
W.H. FREEMAN et Cie, SAN FRANCISCO et LONDRES
(1963).
- [34] S. SOUMARE, J. LOSFELD, APPORTS DE LA TAXONOMIE NUMÉRIQUE A L'ÉTUDE DU
R. BLONDEAU SPECTRE BACTÉRIEN DE LA MICROFLORE DES SOLS DU
NORD DE LA FRANCE,
Ann. Microbiol. (Institut Pasteur), 124 B
pp. 81-94 (1973).
- [35] A. TAQUET, M.T. PLANCOT ÉTUDES PRÉLIMINAIRES SUR LA CLASSIFICATION NUMÉRIQUE
J. DEBRUYNE DES MYCOBACTÉRIES ET DES NOCARDIAS
B. DEVULDER, M. JOSEPH, RELATIONS TAXONOMIQUES ENTRE MYCOBACTERIUM RHODOCROUS
J. LOSFELD, MYCOBACTERIUM PELLEGRINO ET LES GENRES MYCOBACTERIUM
ET NOCARDIA,
Annales de l'Institut Pasteur, LILLE, n° XXII,
pp. 121-135 (1971).
- [36] M. TERRENOIRE UN MODÈLE MATHÉMATIQUE DE PROCESSUS D'INTERROGATION :
LES PSEUDO-QUESTIONNAIRES,
Thèse, GRENOBLE (1970).

[37] R. THOM

STABILITE STRUCTURELLE ET MORPHOGENESE,
W.A. BENJAMIN, MASSACHUSETTS (1972).

[38] N. WIENER

CYBERNETICS,
HERMANN (Act. Sc. 1053), PARIS (1948).

[39] A.M. YAGLOM,
I.M. YAGLOM

PROBABILITE ET INFORMATION,
DUNOD, PARIS (1969).



DEUXIEME THESE

EQUIVALENCE D'EXPRESSIONS RATIONNELLES

Rapporteur M. NIVAT