N° d'ordre : 955

THÈSE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

pour obtenir le titre de

DOCTEUR DE 3eme CYCLE

(Automatique)

par

Victor Manuel TORO CORDOBA

Ingénieur Université des Andes (Bogota-Colombie)

"CONTRIBUTION A L'ANALYSE STRUCTURALE DE SYSTEMES COMPLEXES A L'AIDE DE L'ENTROPIE ET SES GENERALISATIONS"

Thèse soutenue le 8 mars 1982 devant la Commission d'Examen

Membres du Jury :	MM.	P. VIDAL	Président
		M. STAROSWIECKI	Rapporteur
		B. DUBUISSON	Examinateurs
		J. LOSFELD	

- AVANT - PROPOS -

Le travail présenté dans ce mémoire a été effectué au Laboratoire d'Automatique de l'Université des Sciences et Techniques de Lille 1.

Nous remercions Monsieur le Professeur Pierre VIDAL pour l'accueil qu'il nous a réservé dans son Laboratoire et pour l'honneur qu'il nous fait de présider le Jury de cette thèse.

Nos plus vifs remerciements vont à Monsieur Marcel STAROSWIECKI, Maître-Assistant à l'Université de Lille 1, qui nous a suivi et dirigé presque quotidiennement tout au long de ce travail. Ses conseils et orientations, ainsi que son enthousiasme et ses encouragements ont été pour l'essentiel dans les résultats de nos recherches. Qu'il sache que nous garderons toujours un chaleureux souvenir de son amitié.

Nos remerciements vont aussi à Monsieur B.DUBUISSON, Professeur à l'Université de Compiègne, qui a bien voulu juger ce travail.

Que Monsieur J. LOSFELD, Professeur à l'Université de Lille 1, accepte nos remerciements pour sa participation au jury, qui nous honore grandement.

Je tiens à remercier également Madame Michèle LELONG pour la gentillesse et le soin qu'elle a apporté à la mise en page de cette thèse, ainsi qu'à Monsieur HOUZÉ qui a assuré le tirage de ce document.

Nous ne saurions terminer cet avant-propos sans remercier Monsieur B.BARFETY, Directeur du CROUS, Madame CROSS et toutes les personnes du Service d'Accueil des Etudiants Etrangers, qui ont tout fait pour faciliter et rendre agréable notre séjour en France.

Enfin, je tiens à remercier tous les amis que j'ai rencontré dans le Nord; grâce à eux ces années en France ont été pour moi une expérience très enrichissante et positive dont je garderai toujours un agréable souvenir.

Titre : " CONTRIBUTION À L'ANALYSE STRUCTURALE DE SYSTÈMES COMPLEXES A L'AIDE DE L'ENTROPIE ET DE SES GÉNÉRALISATIONS "

Auteur : V.M. TORO

Date de Soutenance : 8 Mars 1982

Résumé :

Cette thèse présente une méthodologie générale permettant au chercheur qui commence l'étude d'un système quelconque de se faire une idée globale de la structure de celui-ci, ainsi que d'orienter et de faciliter les phases suivantes de son étude. La première partie fait une synthèse des principaux résultats de la théorie de l'Information, en mettant l'accent sur leur interprétation intuitive et sur les possibilités de généralisation. La deuxième partie étudie les problèmes de Décomposition en sous-systèmes faiblement couplés, et de détermination des relations qui existent entre les variables.

Mots Clés :

ANALYSE STRUCTURALE; SYSTEMES COMPLEXES; ANALYSE INFORMATIONNELLE DES DONNEES; THEORIE DE L'INFORMATION; MODELISATION.

A Maria Consuelo

" ... Ô mon Dieu, aide-moi à échapper des griffes de ce monde linéaire ... ! "

Pierre de FERMAT (1601-1665)

" In short, the newest study of systems, whether in metal or in flesh, is a branch of communication engineering(*), and its cardinal notions are those of message, amount of disturbance or noise, ..., quantity of information, coding technique and so on.

In such a theory, we deal with systems effectively coupled to the external world, not merely by energy flow, their metabolism, but also by a flow of impressions, of incoming messages, and by de action of outgoing messages ... "

Norbert WIENER
"Cybernetics" . 1ère édition
HERMANN & Cie. Paris. 1948

" ... a theory with mathematical beauty is more likely to be correct than an ugly one that fits some data ... "

P.A.M. DIRAC

" Can equations of physics be used in high energy physics "

Physics Today . April 1970

(*) *Cette "branch of communication engineering" sera appelée plus tard "Théorie de l'Information".*

TABLE DES MATIÈRES

INTRODUCTION

Dans les premières phases de l'analyse d'un système, le praticien a besoin d'outils et de méthodes qui lui permettent de "se faire une idée" globale du système. Ceci est d'autant plus vrai que celui-ci est plus complexe et que la masse de données, de variables et de relations déborde sa capacité d'assimilation.

A ce niveau de l'analyse, les outils et méthodes utilisés doivent posséder au moins les trois caractéristiques suivantes :

- GENERALITE : ils doivent être applicables sans requérir d'hypothèses restrictives, dont le praticien n'est pas en mesure de vérifier la validité.

- GLOBALITE : ils doivent fournir des renseignements globaux qui aident à simplifier la représentation que le praticien se fait du système.

- ORIENTATION : ils doivent permettre d'orienter et de faciliter les phases suivantes de l'étude.

L'Analyse Structurale des Systèmes prétend être une réponse à cette nécessité, et c'est depuis une dizaine d'années que l'on commence à publier des articles qui se réclament de cette approche. On distingue en particulier les travaux de W.Ross ASHBY, Roger C.CONANT, Marc RICHETIN, George J.KLIR, John N.WARFIELD, parmi les initiateurs.

Du fait que les domaines d'application de "l'approche Systèmes" ne cessent de s'élargir (pour inclure aujourd'hui des systèmes socio-économiques, écologiques, biologiques, qui font intervenir des variables non linéaires, non numériques, floues ...) les

techniques de l'Analyse Structurale prétendent avoir une porté plus étendue que d'autres plus "classiques". Ainsi, au lieu des méthodes Statistiques, d'Analyse Numérique, d'Identification, ..., on fait appel à l'Analyse des Données, la Théorie de l'Information, la Théorie des Graphes, la Théorie des Sous-ensembles Flous, la Théorie des Jeux,... .

Il existe aujourd'hui une littérature assez abondante sur le sujet, souvent originale et intéressante, mais à laquelle on peut reprocher son hétérogénéité (de notation, critères d'optimalité, langage, ...). De plus, les recherches ont porté souvent sur la caractérisation et les propriétés des "bonnes structures" d'un système, sur les différents critères d'optimalité, les formes de représentation, ..., sans proposer les méthodes concrètes de recherche ni les algorithmes de construction de ces "bonnes structures"; or le problème, comme on le verra, est loin d'être simple. Enfin, l'utilisation de différentes théories et techniques (Théorie de l'Information, Indices de Ressemblance, Théorie des Graphes,...) est faite d'une manière très indépendante, malgré une certaine similarité des démarches. C'est tout récemment que, dans un remarquable travail, Zohir ABID [Octobre 1979] montre comment ces techniques et théories possèdent un grand nombre de propriétés fondamentales susceptibles d'être harmonisées sous un même formalisme.

Dans ce sens, l'objectif poursuivi dans ce travail est double :

1°) Faire une synthèse des principaux résultats de la théorie de l'Information concernant les systèmes généraux, la plupart connus et quelques uns nouveaux, en mettant un accent tout particulier sur son interprétation intuitive. Ce faisant, et ceci nous semble un point important, nous essayerons de mettre en lumière les hypothèses que sont à la base et sur lesquelles toutes les autres propriétés s'appuient. On isolera alors quatre groupes d'hypothèses : Semi-valuation (SV), Semi-valuation Monotone (SVM), Semi-modulation Positive (SMP), et Semi-modulation Positive Monotone (SMPM). Ainsi, et bien que par souci de clarté les interprétations soient faites en termes de théorie de l'Information, tous les résultats et algorithmes

présentés seront étiquetés par l'un de ces sigles et pourront être généralisés immédiatement et sans aucun changement à d'autres domaines vérifiant les conditions respectives (l'Analyse de Variance, l'Analyse par Indices de Similarité, ...,).

2°) Etudier deux méthodes d'Analyse Structurale : d'une part, la décomposition d'un système en sous-systèmes faiblement couplés; d'autre part, la détermination des interconnexions. Pour chaque méthode, une fois ses caractéristiques étudiées, on présente plusieurs algorithmes détaillés, ainsi que l'analyse de leurs propriétés d'optimalité et de convergence, du temps d'exécution et de la capacité de mémoire requis.

L'application récursive de ces deux méthodes va suggérer une méthodologie complète que nous appellons l'*Analyse Structurale Hiérarchique* : étant donné un système complexe, on commence par trouver la décomposition en sous-systèmes les plus faiblement couplés; on détermine alors les interconnexions entre les sous-systèmes. A nouveau on trouve dans chaque sous-systèmes les sous-sous-systèmes les plus faiblement couplés; puis on détermine les interconnexions entre les sous-sous-systèmes. Et ainsi successivement jusqu'à arriver au niveau le plus élémentaire.

PRESENTATION DES CHAPITRES SUIVANTS

Au CHAPITRE I, nous introduisons une représentation générale des systèmes, qui présente l'avantage de rendre tout à fait analogue, au niveau des formalismes, le cas statique et le cas dynamique. On établit une "*Hiérarchie Epistémologique*" des différents niveaux de connaissance qu'un observateur peut avoir d'un système, et on étudie les structures et propriétés algébriques des niveaux inférieurs de cette hiérarchie. Les différentes démarches de l'Analyse Structurale pourront alors être vues comme des déplacements dans la Hiérarchie Epistémologique.

Aux structures algébriques des niveaux inférieurs le CHAPITRE II vient ajouter l'Entropie et les différents indices de la théorie de l'Information. Une attention spéciale est portée à l'interprétation intuitive de chaque résultat, et au fait que tous les résultats découlent de quelques hypothèses très simples; on verra plus tard que bien d'autres indices issus de domaines différents vérifient aussi ces hypothèses.

Le CHAPITRE III est une étude détaillée d'une première méthode d'Analyse Structurale : la Décomposition en sous-systèmes faiblement couplés. La notion de Décomposition Optimale est discutée et plusieurs algorithmes sont présentés, analysés et comparés.

Dans le CHAPITRE IV on étudie une deuxième méthode d'Analyse Structurale : pour chaque variable X (respec.sous-système S) trouver les autres variables (respec.sous-systèmes) qui expliquent le mieux la variable X(respec.le sous-système S).La relation de dépendance étant la principale raison qui fait qu'une variable ou sous-système explique une autre, cette méthode permet de trouver un graphe des inter-connexions du système. On introduit les concepts de partie expliquée, partie explicable, partie modélisable, partie encore non expliquée de la partie explicable,..., et leurs relations. Plusieurs algorithmes pour la mise en oeuvre de cette méthode sont présentés et analysés. Enfin, on discute brièvement d'autres utilisations possibles : Aide à la représentation et visualisation graphique du système, identification des "équations" qui relient les variables, application à la Reconnaissance des Formes.

CHAPITRE I

HIÉRARCHIE ÉPISTEMOLOGIQUE DES SYSTÈMES

I - INTRODUCTION

Dans un souci de clarification nécessaire à la formulation d'une méthodologie générale, George J.KLIR a proposé et développé une Hiérarchie Epistémologique des Systèmes [KLIR 69,75,76,77] , cherchant à organiser les différents niveaux de connaissance qu'un observateur peut avoir d'un système. Ainsi, par exemple, la connaissance des équations exprimant le comportement de chaque variable en fonction de celui des autres relèvera d'un niveau épistémologique "plus élevé" que, disons, celui de la connaissance de la matrice de covariance entre les variables.

Cette Hiérarchie se traduit par une série de définitions (Système-Source, Système-Données, Système-Générateur, Système-Structure, Méta-Système, ...) plus ou moins emboîtées les unes dans les autres. Reprenant l'idée générale de G.J. KLIR, nous essayons d'y ajouter trois aspects :

. Nous proposons une forme de représentation générale, qui s'adapte à tous les types de systèmes, et qui offre l'avantage d'unifier le traitement des Systèmes Statiques et Dynamiques. En effet, très souvent les cas statique et dynamique sont traités d'une façon tout à fait différente, et les algorithmes et méthodes applicables pour un cas n'ont pas une équivalence claire pour l'autre.

. Les manipulations et démarches faites sur les systèmes seront, elles aussi, hiérarchisées au niveau épistémologique nécessaire à leur application. Par exemple, si X est une variable d'un système, il est facile d'accepter qu'une démarche visant à minimiser la fonction C(X) requiert un niveau de connaissances sur le système "plus élevé" que si l'on cherchait tout simplement à "estimer la distribution de probabilité de la variable X".

. Dans chaque niveau de la hiérarchie épistémologique, en particulier au niveau Système-Donnée, nous essayons d'expliciter les concepts et propriétés fondamentales avec un certain détail. Cette démarche, qui pourrait paraître superflue ou triviale, cherche à mettre en lumière, les conditions et les propriétés sur lesquelles toutes les autres s'appuient, afin de rendre possible leur généralisation ultérieure.

2 - APERCU DE LA HIÉRARCHIE ÉPISTÉMOLOGIQUE DES SYSTÈMES

Un système au niveau épistémologique le plus bas, le Niveau 0, est appelé Système-Source . Il est défini par un ensemble de variables et un ensemble des *Modalités* ou états potentiels associé à chaque variable. Aucune connaissance sur la probabilité des différentes modalités ou sur l'évolution des variables n'est supposée à ce niveau. Aux niveaux épistémologiques supérieurs on disposera d'avantage de connaissances d'abord sur les variables, et ultérieurement, sur les relations entre elles.

Au Niveau 1 on parlera de Système-Donnée . Le système est connu par une succession d'états ou modalités prises par les variables. Ceux-ci peuvent être réels (dans le cas d'un échantillonnage d'un système réel), simulés (si un modèle existe déjà), ou désirés (dans le cas d'une étude de conception). Le système se présente alors comme un *Tableau Initial de Données* où chaque ligne est constituée des modalités que les variables prennent simultanément. (Désormais nous utilisons le mot *modalité* au lieu de *état*, ce dernier étant en général compris comme une valeur numérique, ce qui n'est pas forcément le cas dans ce travail).

Les niveaux supérieurs ajoutent la connaissance de quelques propriétés ou relations *Invariantes* (au moins localement) à travers lesquelles le tableau des données pourrait être obtenu ou approximé de façon déterministe ou stochastique, pour des conditions initiales appropriées.

Au niveau 2 le système est défini par un ensemble de *Relations Génératrices* invariantes permettant de connaître (exactement ou approximativement) le comportement de quelques variables en fonction de celui des autres. Ces relations génératrices peuvent aller des distributions de probabilités conjointes (permettant de construire approximativement un tableau de données par simulation), jusqu'à des équations liant différentes variables. Ces relations peuvent aussi contenir d'autres variables ne faisant pas partie du système : variables d'entrée, ou tout simplement, variables de l'environnement. On parlera à ce niveau de Système-Générateur.

Au niveau 3 un système est composé d'un ensemble de *Sous-Systèmes Générateurs*, plus certaines relations ou couplages entre eux. Il s'agira à ce niveau de Systèmes-Structure. Différents types de structures peuvent être envisagés : au chapitre III nous étudierons le cas ou les sous-ensembles de variables correspondantes aux sous-systèmes générateurs forment une partition; au chapitre IV ce sera une pseudo-partition.

G.J.KLIR propose encore des niveaux supérieurs :
le Méta-Système, constitué par un ensemble de Sous-Systèmes-Structure;
le Méta-Méta-Système, formé par un ensemble de Sous-Méta-Systèmes,
etc ..., mais leur considération dépasse le cadre de ce travail.

Bien sur, cette hiérarchie (tout au moins la description que nous en donnons ici) ne prétend pas être très rigoureuse, mais seulement distinguer un peu les niveaux de discussion sur un système. Il est par ailleurs possible que les connaissances que l'on possède sur un système ne permettent pas de s'identifier tout à fait à un niveau épistémologique, soit parce qu'elles sont incomplètes (par exemple lorsque on connait les variables contenues dans chaque sous-système générateur, mais on ignore les relations génératrices),soit parce qu'elles relèvent de plusieurs niveaux (on connait les variables de chaque sous-système générateur et le sous-tableau de données respectif). Nous reviendrons plus en détail sur cette hiérarchie au paragraphe 4.

3 - L'ANALYSE STRUCTURALE ET LA HIERARCHIE ÉPISTÉMOLOGIQUE

La Hiérarchie Epistémologique étant définie, on se pose tout de suite le problème du déplacement entre ses différents niveaux. Tout d'abord, celui de descendre dans la hiérarchie, qui ne présente en principe aucune difficulté puisque par définition la connaissance d'un niveau implique celle du niveau inférieur. Par contre, celui de monter dans la hiérarchie semble moins simple (quelle que soit la hiérarchie, d'ailleurs !). On peut considérer deux cas :

- Du niveau 1 au niveau 2 : Le système est connu au moyen d'un tableau de Données et on cherche les relations génératrices qui permettent d'approximer au mieux ces données. Pour le cas des systèmes à variables numériques (plus d'autres hypothèses de continuité, linéarité, etc ...), ce problème relève de la Modélisation et de l'Identification. Pour des conditions plus générales une discussion est présentée dans [KLIR 75].

- Vers le niveau 3 : A partir du niveau 1 ou 2, on peut chercher une représentation du type Système-Structure. *Cette démarche est justement celle de l'Analyse Structurale*. On peut distinguer trois cas :

3.1 - ANALYSE STRUCTURALE MATHEMATIQUE (Niveau 2 → Niveau 3)

Nous appelons ainsi l'ensemble des techniques et méthodes qui, à partir des relations génératrices du système (en général des équations), cherchent une représentation structure que s'y adapte. Une telle démarche s'avère fort utile pour la simplification et résolution des systèmes décrits par de nombreuses équations [STEW 65; GABA 78; NEPO 78; RICH 75, chap.II]. Mais ainsi que le remarque J.DUFOUR [1979, chp.I], outre le fait que les équations sont supposées connues (ce qui n'est pas le cas dans les premières phases de l'étude d'un système), les hypothèses requises sont quelquefois

nombreuses et les résultats obtenus peuvent s'éloigner sensiblement de la structure "naturelle", au point qu'il peut être impossible d'associer une signification physique aux sous-systèmes (ou sous-structures) dégagés.

3.2 - ANALYSE STRUCTURALE PHYSIQUE (? → Niveau 3)

Il s'agit d'une approche "naturelle" à laquelle il est difficile d'atrribuer un niveau épistémologique de départ. Le praticien, guidé par son expérience, par les objectifs du système, par certaines contraintes, etc ..., est amené à distinguer très directement et très concrètement des ensembles d'éléments qui sont connectés plus ou moins fortement par des flux de matières, d'énergie, d'information, ou même par de critères tels que même technologie, même fournisseur, type de personnel qui y travaille, etc ... . Ce type d'analyse, guidé avant tout par le "bon sens", est sans doute celui que l'on retrouve le plus souvent. Il permet de dégager les très grandes lignes de la structure et uniquement celles-ci, mais il se révèle incapable de procéder à des analyses plus fines du système [DUFO 79, chap.I].

3.3 - ANALYSE STRUCTURALE PAR LES DONNEES (Niveau 1 → Niveau 3)

A partir d'un Tableau Initial des Données supposé être *la seule connaissance disponible du système* on cherche une (des) représentation stucture aussi appropriée que possible. L'analyse s'appuie sur l'hypothèse que cette structure est toujours décelable dans les données produites par le système, et qu'elle se concrètise par des évolutions non indépendantes de certains groupes de variables [SIMO 62]. A l'aide de la théorie de l'Information et de l'Analyse de Données nous étudierons au chapitre III les représentations structure où ces

groupes de variables dépendantes forment une partition, et au chapitre IV nous utiliserons un type particulier de pseudo-partition.

L'Analyse Structurale par les Données se situe à mi-chemin entre les Analyses Structurales Physique et Mathématique [DUFO 79, chap.I] . En effet :

- d'une part elle utilise les informations recueillies sur le système et donc reste toujours liée à ce système, à ses constituants, à leur structure interne.

- d'autre part elle s'appuie sur des méthodes mathématiques ayant des bases statistiques et/ou informationnelles, et s'écarte ainsi d'une spéculation purement pragmatique.

4 - VERS UNE REPRÉSENTATION DU TYPE SYSTÈME-STRUCTURE

Après avoir présenté d'une façon globale la Hiérarchie Epistémologique des Systèmes, nous la reprenons niveau par niveau dans l'optique d'une démarche " *vers le niveau 3* ", celui des Systèmes-Structure. Nous introduisons à chaque niveau la notation et les concepts algébriques nécessaires.

4.1 - SYSTEMES-SOURCE (Niveau 1)

Soit $\{ X_1, X_2, X_3, \ldots, X_N \}$ l'ensemble des N *variables Primaires* que le chercheur considère importantes pour la description du système étudié. Ce sont des variables intervenant dans le système (variables d'entrée, de sortie, de commande, d'état), ou caractérisant son fonctionnement ou son observation [RICH 75, chap.III] . Le choix de ces variables dépend avant tout de l'expérience du chercheur et des connaissances qu'il possède déjà sur le système (au chapitre IV, nous proposons quelques mesures du "bon choix" de cet ensemble de variables primaires).

On considère que l'ensemble des variables primaires est *ordonné* (un ordre quelconque mais fixe), pour éviter des ambiguités lorsque on sera amené à considérer cet ensemble lui-même comme une variable vectorielle à N composantes. Il en sera de même pour ses sous-ensembles.

Soient M_1, M_2, ..., M_N les N ensembles finis des *Modalités* prises par les variables X_1, X_2, ..., X_N respectivement, avec $M_i = \{ a_1^i, a_2^i \ldots, a_{m_i}^i \}$. Aucune hypothèse algébrique n'est faite ni sur les variables X_i ni sur les ensembles de modalités M_i, et il peut même s'agir de variables décrivant des caractéristiques non quantitatives du système. Ceci est souvent le cas des systèmes socio-économiques, pour lesquels il importe de respecter la représentation naturelle sans leur imposer des structures particulières [LERM 76, 81 Introduc].

4.2 - SYSTEME-DONNEES (Niveau 1)

Le relevé des modalités effectivement prises par les variables décrivant le système dépend du caractère statique ou dynamique de celui-ci :

- pour le cas statique, c'est-à-dire lorsque l'évolution du système est nulle ou si lente qu'il n'est pas pratique de l'étudier en observant l'évolution de ses variables, on est amené à considérer une population de L répliques ou exemplaires du système et à étudier les modalités prises par les variables sur chacun des individus de cette population. C'est le cas typique des sondages d'opinion. L'ordre des individus dans la population choisie est alors indifférent.

- pour le cas dynamique, nous allons toujours considérer les systèmes sous une forme échantillonnée et discrète. Cette limitation obéit à des raisons pratiques (le traitement en ordinateur digital implique forcément un échantillonnage dans le temps et une discrétisation de l'espace des modalités des variables) et mathématiques (l'entropie et les autres indices considérés ne possèdent pas sur les variables continues toutes les "bonnes propriétés" qu'ils vérifient sur les variables discrètes). Ainsi, nous allons supposer que l'on observe les variables aux instants t_0, $t_0 + \Delta t$, $t_0 + 2\Delta t, \ldots$, $t_0 + (L-1)\Delta t$. Bien sur, le choix de t_0, Δt et L est important; il sera discuté au paragraphe 4.3.1. .

Les données relevées se présentent alors sous la forme d'un *Tableau Initial de Données* (Fig.I.1).

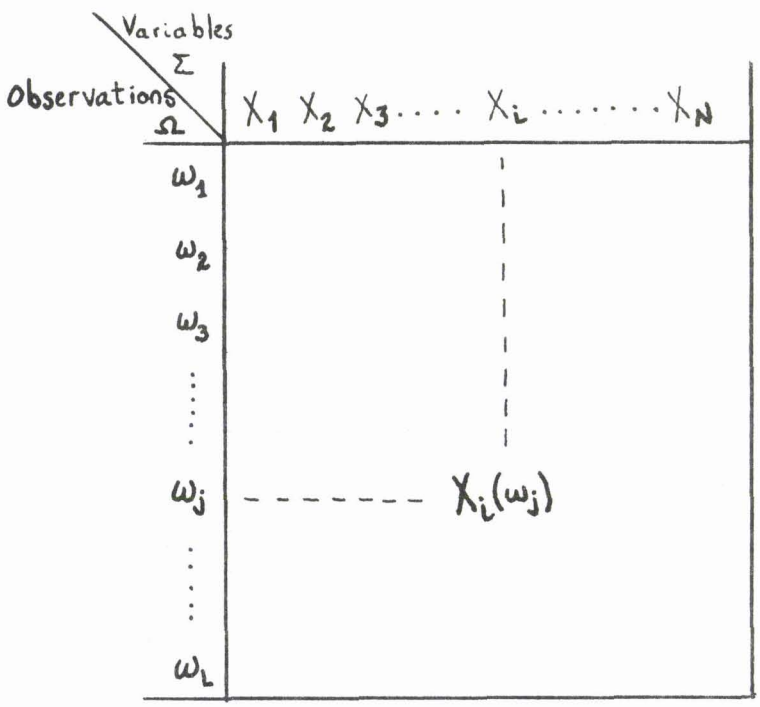


Figure I.1: Tableau initial de données

où ω_j représente soit l'individu j d'une population de taille L, soit le système à l'instant $t_0 + (j-1)\cdot\Delta t$. La ligne j contient les modalités prises par chacune des variables sur ω_j. Pour un bon échantillonnage, la colonne i devrait contenir tous les éléments de M_i. *Ce tableau Initial de Données constitue la seule connaissance de départ pour toute notre analyse.*

Nous donnons ci-dessous quelques définitions formelles séparément pour les systèmes statiques et les systèmes dynamiques.

4.2.1. - Représentation des Systèmes Statiques (E)

Voici les éléments qui constituent un Système-Donnée Statique :

- $\Omega^E := \{ \omega_1, \omega_2, \ldots, \omega_L \}$ est l'*Ensemble d'Echantillonnage Statique.*

- $X_i^E : \Omega^E \longrightarrow M_i^E$ $\qquad i=1,2,\ldots,N$

$\omega_j \longmapsto X_i^E(\omega_j) := X_i(\omega_j)$

sont les *Variables du Système Statique.* $M_i^E := M_i = \{a_1^i, a_2^i, \ldots, a_{m_i}^i\}$ est l'*Ensemble des Modalités de la Variable* X_i^E.

- $\Sigma^E := \{X_1^E, X_2^E, \ldots, X_N^E\}$ est l'*Ensemble Ordonné des Variables du Système Statique.* Le fait de le considérer ordonné permet d'avoir une écriture unique pour Σ^E et chacun de ses sous-ensembles.

- Le *Tableau de Données Statique (TDE)* est montré dans la figure I.2 :

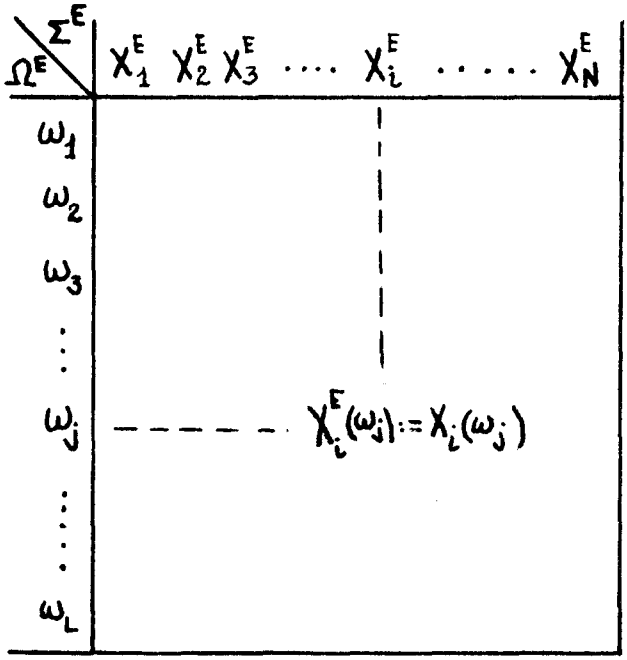


Figure I.2 : Tableau des Données Statique (TDE)

4.2.2. - Représentation des Systèmes Dynamiques (D)

Dans les systèmes dynamiques, en général, la détermination d'une variable à un instant donné requiert l'observation préalable de l'évolution du système pendant un certain nombre d'intervalles de temps. L'*Ordre du Système* est défini justement comme ce nombre p d'observations préalables nécessaires à la détermination de l'évolution future du système. Pour chaque variable primaire X_i, on définit alors une variable d'état X_i^D dont la valeur à chaque instant est un vecteur qui a comme composantes les p dernières valeurs prises par la variable primaire X_i.

Cependant, avec les connaissances que nous possédons sur le système à ce niveau, il est impossible de déterminer avec certitude l'ordre du système. Nous supposons que le praticien choisit un ordre p basé sur son expérience, sur la taille du tableau initial des données, et sur la capacité de l'ordinateur sur lequel les algorithmes qui seront décrits vont être exécutés. Au chapitre IV, nous donnerons quelques indices permettant de qualifier le choix de p.

Cela dit, voici les éléments formels qui constituent un Système-Donnée Dynamique :

- $\Omega^D := \{ \omega_p, \omega_{p+1}, \ldots, \omega_L \}$ est l'*Ensemble d'Echantillonnage Dynamique.*

On peut remarquer, que par rapport au cas statique, les p-1 premiers éléments ont été éliminés afin de pouvoir considérer dans tous les cas les p dernières réalisations d'une variable.

$$- X_i^D : \Omega^D \longrightarrow M_i^D \qquad i=1,2,\ldots,N. \qquad [I.1]$$

$$\omega_j \longmapsto X_i^D(\omega_j) := (X_i(\omega_j), X_i(\omega_{j-1}), X_i(\omega_{j-2}), \ldots, X_i(\omega_{j-(p-1)}))$$

sont les *variables du Système Dynamique.* $M_i^D := M_i^p = M_i \times M_i \times \ldots \times M_i$ (p fois) est l'*Ensemble de Modalités de la variable* X_i^D . Ces variables sont donc des vecteurs à p composantes.

- $\Sigma^D := \{X_1^D, X_2^D, \ldots, X_N^D\}$ est l'*Ensemble Ordonné des N variables du Système Dynamique.* Par les mêmes raisons que pour le cas statique, on considère Σ^D comme étant ordonné.

- Le *Tableau de Données Dynamique (TDD)* est montré dans la figure I.3

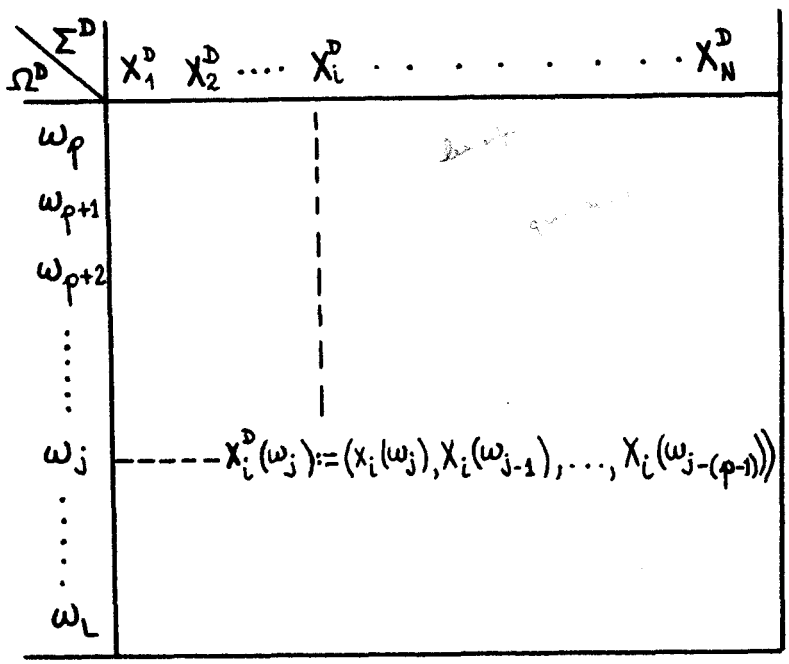


Figure I.3 : Tableau des Données Dynamique (TDD)

Il faut remarquer que chaque élément du TDD est un vecteur de p composantes. Mais en réalité le TDD n'est qu'un formalisme mathématique, puisque dans la pratique on conserve le même tableau Initial de Données, la prise en compte à chaque instant des p dernières valeurs des variables se faisant au niveau des algorithmes.

Comme on le voit immédiatement, les représentations E (pour statique) et D (pour dynamique) sont tout à fait analogues. Désormais donc, *chaque fois que l'on parle de* Ω, X_i, M_i, Σ *ou du Tableau de Données, sans spécifier ni E ni D, il pourra s'agir aussi bien de* Ω^E, X_i^E, M_i^E, Σ^E *et du* TDE, *ou bien de* Ω^D, X_i^D, M_i^D, Σ^D *et du* TDD.

4.2.3. - Le Inf-Demi-Treillis des Partitions de Ω

Nous allons maintenant présenter brièvement quelques concepts bien connus de la Théorie des Ensembles. Pour cela, il est peut être convenable d'imaginer Ω, X_i, M_i et Σ d'un point de vue plus abstrait: Σ est tout simplement un ensemble de N variables $X_i : \Omega \longrightarrow M_i$, où Ω et M_i sont eux aussi simplement des ensembles. Ceci peut être visualisé dans la figure I.4 .

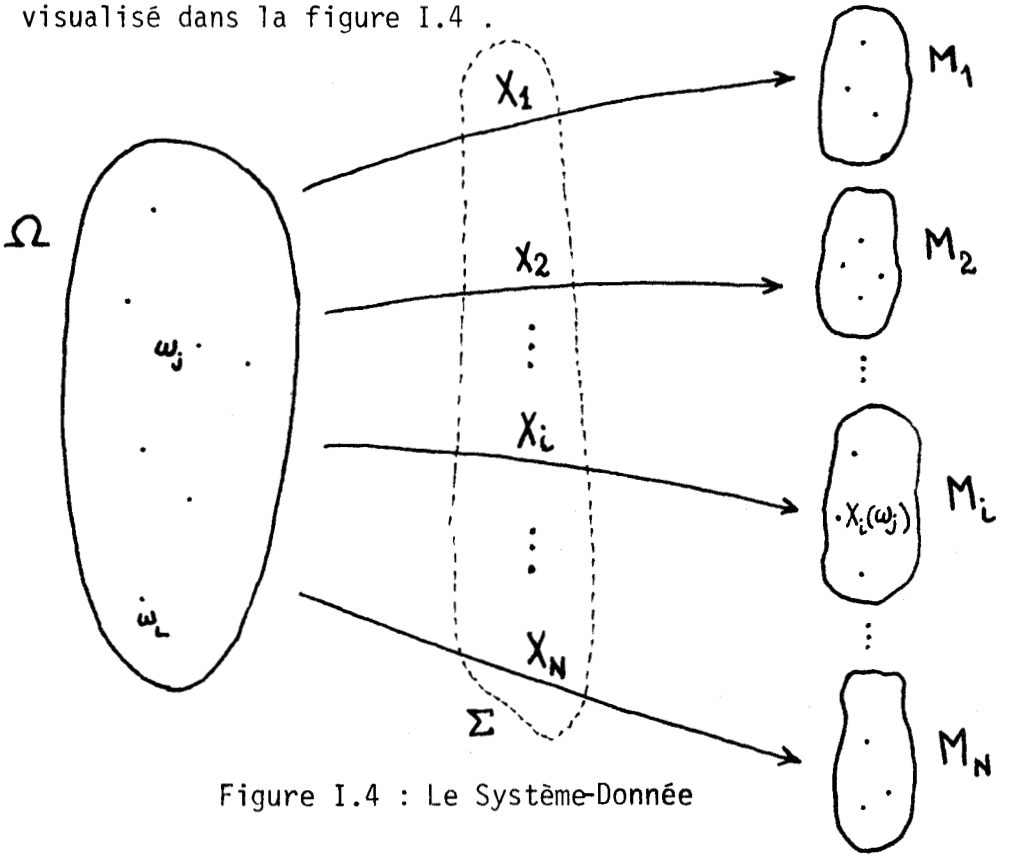


Figure I.4 : Le Système-Donnée

Soit $\mathcal{P}(\Omega) := \{ C : C \subseteq \Omega \}$ l'ensemble de tous les sous-ensembles de Ω.

On dit que un ensemble $F = \{ F_1, F_2, \ldots, F_k \}$ est une *Partition de Ω en k classes* ssi :

- $F_i \in \mathcal{P}(\Omega)$ et $F_i \neq \emptyset$ pour $i = 1,2,\ldots, k$; et

- $\forall i \neq j : F_i \cap F_j = \emptyset$; et

- $\bigcup_{i=1}^{k} F_i = \Omega$.

Soit $\mathbb{P}(\Omega)$: = {F:F est une partition de Ω},l'*ensemble de toutes les partitions de* Ω. Dans cet ensemble on peut établir la relation $\preccurlyeq$:*"être plus fine que "*, définie ainsi : si $F = \{F_1,F_2,\ldots,F_s\}$ et $G = \{G_1,G_2,\ldots,G_r\}$ sont deux éléments de $\mathbb{P}(\Omega)$,

$$F \preccurlyeq G :\Leftrightarrow \quad \forall F_i \in F , \exists G_j \in G : F_i \subseteq G_j \qquad [I.2]$$

Par ailleurs, si $F = \{F_1,F_2,\ldots F_s\} \preccurlyeq G = \{G_1,G_2,\ldots,G_r\}$, chaque classe $G_j \in G$ est justement l'union des classes F_i qu'elle contient ($\forall G_j \in G : G_j = \bigcup_{i:F_i \subseteq G_j} F_i$). Moyennant une renumérotation, cela peut s'écrire :

$$F = \{ \underbrace{F_1,F_2,\ldots,F_a}, \underbrace{F_{a+1},\ldots, F_b}, \underbrace{F_{b+1},\ldots\ldots,F_u}, \ldots, F_s \}$$

$$G = \{ \quad G_1, \qquad G_2, \qquad G_3,\ldots, \quad G_r \quad \}$$

Ainsi, G détermine une partition des classes de F.

Cette relation $\preccurlyeq$ est un ordre partiel, c'est-à-dire qu'elle vérifie les trois conditions :

- $F \preccurlyeq F \qquad \forall F \in \mathbb{P}(\Omega)$ (Reflexive)

- $F \preccurlyeq G$ et $G \preccurlyeq F \Rightarrow F = G$ (Anti-symétrique)

- $F \preccurlyeq G$ et $G \preccurlyeq D \Rightarrow F \preccurlyeq D$ (Transitive).

$\mathbb{P}(\Omega)$ muni de la relation $\preccurlyeq$ constitue un *Treillis* : n'importe quel couple F,G d'élément de $\mathbb{P}(\Omega)$ possède un *Infimum* (noté $F \wedge G$) et un *Supremum* (noté $F \vee G$) définis ainsi :

- $(F \wedge G) \preccurlyeq F$ et $(F \wedge G) \preccurlyeq G$, et $\forall D$ t.q. $D \preccurlyeq F$ et $D \preccurlyeq G$, alors $D \preccurlyeq (F \wedge G)$ [I.3]

- $(F \vee G) \succcurlyeq F$ et $(F \vee G) \succcurlyeq G$, et $\forall D$ t.q. $D \succcurlyeq F$ et $D \succcurlyeq G$, alors $D \succcurlyeq (F \vee G)$ [1.4]

$F = \{F_1, F_2, \ldots, F_s\}$ et $G = \{G_1, G_2, \ldots, G_r\}$ étant deux partitions de Ω, la partition $F \wedge G$ est donnée par

$$F \wedge G = \{F_i \cap G_j : i=1,2,\ldots,s \ , \ j=1,2,\ldots,r, \text{ et } F_i \cap G_j \neq \emptyset \} \qquad [I.5]$$

$$= \{F_1 \cap G_1, F_1 \cap G_2, \ldots, F_1 \cap G_r, F_2 \cap G_1, \ldots, F_2 \cap G_r, \ldots, F_s \cap G_r\} \setminus \{\emptyset\}$$

(pour l'obtention de $F \vee G$, plus compliquée à décrire, voir [LERM 76,81,chap.0]).

La relation entre $\preccurlyeq$ et les opérations $\wedge$ et $\vee$ peut aussi être établie par les équivalences suivantes [BIRK 64] :

$$F \preccurlyeq G \Longleftrightarrow F \wedge G = F \Longleftrightarrow F \vee G = G \qquad [I.6]$$

Nous ne retiendrons que quelques unes des propriétés du Treillis $< \mathbb{P}(\Omega), \leqslant, \wedge, \vee >$, résumées dans la proposition suivante :

Proposition I.1

Etant donné un ensemble fini Ω, l'ensemble $\mathbb{P}(\Omega)$ de toutes les partitions de Ω, muni de l'opération $\wedge$:

$$\wedge : \mathbb{P}(\Omega) \times \mathbb{P}(\Omega) \longrightarrow \mathbb{P}(\Omega)$$
$$(F,G) \longmapsto F \wedge G \text{ définie par } [I.3] \text{ ou } [I.5]$$

constitue un *Inf-Demi-Treillis*, c'est-à-dire, vérifie les conditions suivantes, quelques soient $F,G,D \in \mathbb{P}(\Omega)$:

i.	$F \wedge F = F$	(idempotence)
ii.	$F \wedge G = G \wedge F$	(commutativité)
iii.	$F \wedge (G \wedge D) = (F \wedge G) \wedge D$	(associativité)

Démonstration

Si on utilise la représentation de $\wedge$ donnée par [I.5], les conditions i,ii,iii découlent immédiatement des propriétés élémentaires de l'intersection. (F.D.)

Un exemple : Pour $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ on donne l'ensemble $\mathbb{P}(\Omega)$ (fig.I.5). Une flèche entre deux partitions indique que celle de départ est plus fine que celle d'arrivée (les flèches issues de la réflexivité et de la transitivité de $\leqslant$ ne sont pas dessinées).

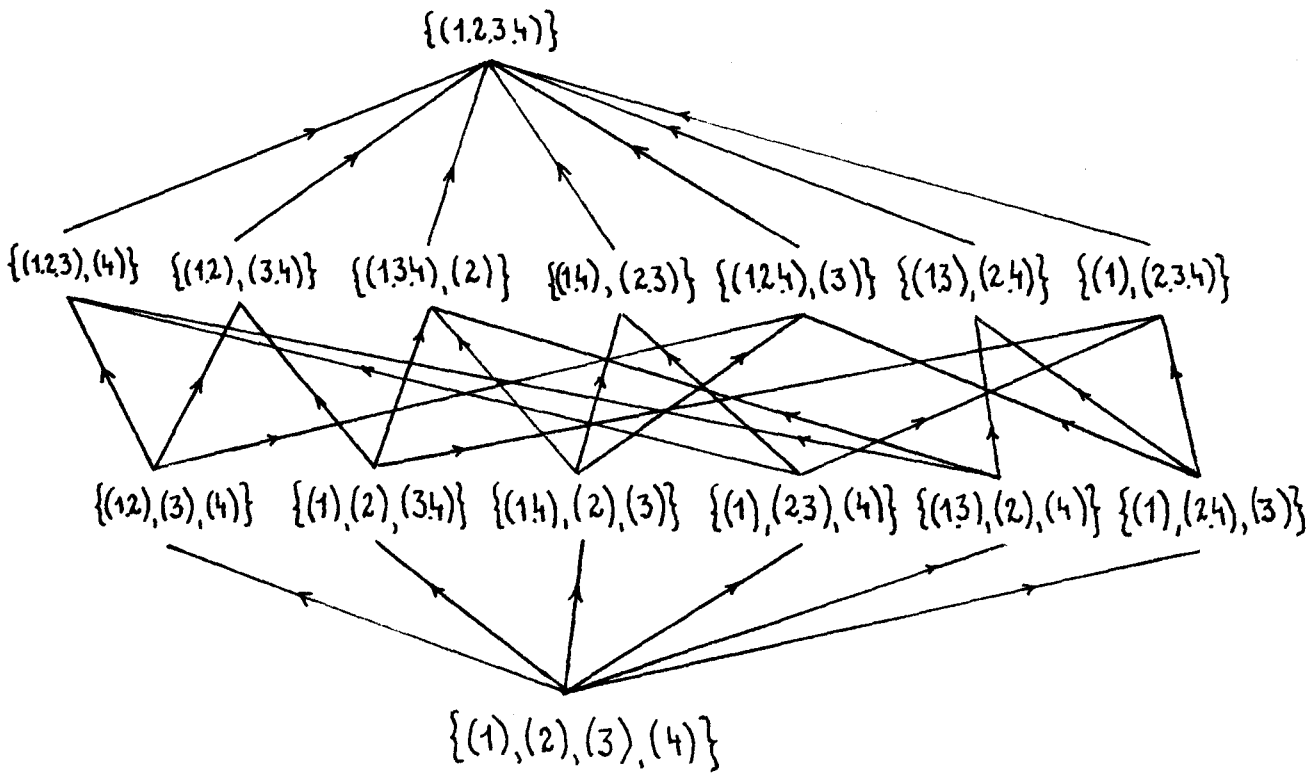


Figure I.5 : Partition de $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4\}$

Il est évident que l'ensemble $\mathbb{P}(\Omega)$ ne dépend que du cardinal de Ω. Ainsi, nous étudierons au chapitre III l'ensemble $\mathbb{P}(\Sigma)$, pour lequel on aura tout à fait les mêmes propriétés que pour $\mathbb{P}(\Omega)$.

4.2.4. - Pourquoi s'intéresse-t-on aux partitions ?

La raison de notre intérêt pour $\mathbb{P}(\Omega)$ est la suivante :

Quelle que soit la variable $X : \Omega \longrightarrow M_X$ on peut définir la *Partition* P_X *de* Ω *associée à* X :

$$P_X := \{X^{-1}(\alpha) : \alpha \in M_X\} \setminus \{\emptyset\} = \{\{\omega : X(\omega) = \alpha\} : \alpha \in M_X\} \setminus \{\emptyset\} \qquad [I.7]$$

c'est-à-dire, la partition de Ω obtenue en regroupant dans une même classe tous les ω auxquels X attribue une même modalité (le terme " $\setminus\{\emptyset\}$ " élimine les $\emptyset$ qui apparaissent lorsque aucun ω ne prend une certaine modalité α). Il faut bien remarquer que la partition P_X *ne dépend pas des modalités* α *en elles mêmes*, mais seulement du fait qu'elles sont différentes. Pour illustrer la dualité statique-dynamique de la notation, voyons la définition [I.7] pour les variables de Σ dans chacun des cas :

- Cas statique : Si $X_i^E : \Omega^E \longrightarrow M_i^E$

$$P_{X_i^E} := \{X_i^{E^{-1}}(\alpha) : \alpha \in M_i^E\} \setminus \{\emptyset\} = \{\{\omega : X_i^E(\omega) = \alpha\} : \alpha \in M_i^E\} \setminus \{\emptyset\} \qquad [I.8]$$

- Cas dynamique : Si $X_i^D : \Omega^D \longrightarrow M_i^D$,

$$P_{X_i^D} := \{X_i^{D^{-1}}(\alpha) : \alpha \in M_i^D\} \setminus \{\emptyset\} = \{\{\omega : X_i^D(\omega) = \alpha\} : \alpha \in M_i^D\} \setminus \{\emptyset\} \qquad [I.9]$$

A travers cette association P entre la variable $X_i \in \Sigma$ et la partition $P_{X_i} \in \mathbb{P}(\Omega)$, quel que soit i=1,2,...,N, on peut dire que $\Sigma \subseteq \mathbb{P}(\Omega)$ (donc $\Sigma^E \subseteq \mathbb{P}(\Omega^E)$ et $\Sigma^D \subseteq \mathbb{P}(\Omega^D)$).

Cette possibilité d'associer univoquement à chaque variable une partition nous permet d'établir un ordre et une équivalence entre variables. Pour cela, si

$$X : \Omega \longrightarrow M_X \qquad\qquad Y : \Omega \longrightarrow M_Y$$
$$\omega \longmapsto X(\omega) \qquad\qquad \omega \longmapsto Y(\omega)$$

sont deux variables quelconques définies sur Ω, on dit que

- X *est équivalente à* Y $:\Leftrightarrow P_X = P_Y$ [I.10]

Dans ce cas il est facile de montrer qu'il existe une fonction $f : M_X \longrightarrow M_Y$ bijective telle que $Y = f \circ X$ et $X = f^{-1} \circ Y$, donc, X ne serait qu'un recodage univoque de Y, et viceversa. Cette affirmation est aussi valable dans l'autre sens : si une telle fonction f existe, alors $P_X = P_Y$, donc X est équivalente à Y.

- X *est plus fine que* Y $:\Leftrightarrow P_X \leqslant P_Y$ [I.11]

Dans ce cas, on peut montrer qu'il existe une fonction $f : M_X \longrightarrow M_Y$ telle que $Y = f \circ X$. Cette affirmation est aussi valable dans l'autre sens : si une telle fonction f existe, alors $P_X \leqslant P_Y$, donc X est plus fine que Y. Cela implique par ailleurs que pour n'importe quelle fonction g définie sur M_X on a toujours que X est plus fine que g (X).

Nous verrons au prochain chapitre que la théorie de l'Information ne distingue pas entre variables équivalentes. C'est pour cela que lorsqu'on fait une analyse des variables d'un système basée sur la théorie de l'Information, $\mathbb{P}(\Omega)$ peutêtre interprété comme l'*ensemble de toutes les variables définies sur* Ω.

Nous allons maintenant étudier un sous-ensemble particulier de $\mathbb{P}(\Omega)$: celui des variables vectorielles dont les composantes sont des variables de Σ.

4.2.5. - Le Sup-Demi-Treillis des Variables Vectorielles

On définit :

$\mathcal{P}(\Sigma) := \{S : S \subseteq \Sigma\}$, l'ensemble de tous les sous-ensembles de Σ;

$\mathcal{P}^*(\Sigma) := \{S : S \subseteq \Sigma \text{ et } S \neq \emptyset\}$, l'ensemble de tous les sous-ensembles non vides de Σ.

Cet ensemble $\mathcal{P}^*(\Sigma)$ peut être vu comme l'ensemble des variables vectorielles dont les composantes sont des variables de Σ. Cela s'explique par la raison suivante : d'une part, deux variables vectorielles qui comportent les mêmes composantes mais en ordre différent sont des variables équivalentes (elles donnent lieu à la même partition de Ω). D'autre part, le fait d'avoir supposé Σ comme un ensemble ordonné fait que chaque élément de $\mathcal{P}^*(\Sigma)$ a une écriture unique. Ainsi, chaque élément $S \in \mathcal{P}^*(\Sigma)$ représente toutes les variables vectorielles qui lui sont équivalentes par permutation des composantes de S.

Détaillons un peu la notation de ces variables vectorielles

Soit $S = \{ X_{i_1}, X_{i_2}, \ldots, X_{i_k} \} \in \mathcal{P}^*(\Sigma) \qquad (i_1 < i_2 < \ldots < i_k)$

On définit :

$$- J_S := \{i : X_i \in S\} = \{i_1, i_2, \ldots, i_k\} \qquad [I.12]$$

l'ensemble ordonné des sous-indices des variables appartenant à S.

$$- M_S := \prod_{i \in J_S} M_i = M_{i_1} \times M_{i_2} \times \ldots \times M_{i_k} \qquad [I.13]$$

l'ensemble de modalités de la variable vectorielle S. Alors, la variable S est définie formellement ainsi :

$$- S : \Omega \longrightarrow M_S \qquad [I.14]$$

$$\omega \longmapsto S(\omega) := (X_{i_1}(\omega), X_{i_2}(\omega), \ldots, X_{i_k}(\omega)).$$

Pour le cas Dynamique, par exemple, les définitions ci-dessus donneraient :

$$\text{Soit } S^D = \{X^D_{i_1}, X^D_{i_2}, \ldots, X^D_{i_k}\} \in \mathbb{P}(\Sigma^D).$$

$$- J_{S^D} := \{i : X^D_i \in S^D\} = \{i_1, i_2, \ldots, i_k\} \qquad [I.15]$$

$$- M_{S^D} := \prod_{i \in J_{S^D}} M_i^D = M_{i_1}^D \times M_{i_2}^D \times \ldots \times M_{i_k}^D \qquad [I.16]$$

$$- S^D : \Omega \longrightarrow M_{S^D}$$

$$\omega \longmapsto S^D(\omega) := (X_{i_1}^D(\omega), X_{i_2}^D(\omega), \ldots, X_{i_k}^D(\omega)) \qquad [I.17]$$

Les définitions [I.12, 13, 14] permettent d'associer une variable vectorielle à chaque $S \in \mathcal{P}(\Sigma)$, $S \neq \emptyset$. À $S = \emptyset$ on peut associer la "variable" constante Φ (il n'y a qu'une seule variable constante puisque elles sont toutes équivalentes) définie ainsi :

$$- \Phi : \Omega \longrightarrow M_\Phi := \{*\} \qquad [I.18]$$

$$\omega \longmapsto \Phi(\omega) := *$$

Bien entendu, $P_\Phi = \{\Omega\}$. (Il y aura donc Φ^E et M_{Φ^E} pour le cas statique, Φ^D et M_{Φ^D} pour le cas dynamique).

Ainsi donc, $\mathcal{P}(\Sigma)$ est l'*ensemble de toutes les variables vectorielles définies sur* Ω (y compris la variable vectorielle constante) *dont les composantes appartiennent à* Σ. Par ailleurs, et du fait que à chaque variable on peut assigner une partition dans $\mathbb{P}(\Omega)$, on peut dire aussi que $\mathcal{P}(\Sigma) \subseteq \mathbb{P}(\Omega)$.

Dans l'ensemble $\mathcal{P}(\Sigma)$, la relation d'inclusion $\subseteq$ définit un ordre partiel.

- $A \subseteq A$, $\forall A \in \mathcal{P}(\Sigma)$ (reflexivité)
- $A \subseteq B$ et $B \subseteq A \Rightarrow A = B$ (anti-symétrie)
- $A \subseteq B$ et $B \subseteq C \Rightarrow A \subseteq C$ (transitive)

De plus, pour tout couple A,B d'éléments de $\mathcal{P}(\Sigma)$, il existe un Infimum (qui correspond à $A \cap B$) et un Supremum (qui correspond à $A \cup B$) tels que :

$(A \cap B) \subseteq A$ et $(A \cap B) \subseteq B$, et $\forall$ D t.q. $D \subseteq A$ et $D \subseteq B \Rightarrow D \subseteq (A \cap B)$ [I.19]

$(A \cup B) \supseteq A$ et $(A \cup B) \supseteq B$, et $\forall$ D t.q. $D \supseteq A$ et $D \supseteq B \Rightarrow D \supseteq (A \cup B)$ [I.20]

La relation entre $\subseteq$, $\cup$ et $\cap$ est donnée par :

$$A \subseteq B \Leftrightarrow A \cup B = B \Leftrightarrow A \cap B = A \qquad [I.21]$$

Ainsi, l'ensemble $\mathcal{P}(\Sigma)$ avec la relation $\subseteq$ et les opérateurs $\cup$ et $\cap$ devient un treillis (ou plus exactement une algèbre de Boole [BIRK 64]). Mais nous retenons seulement une structure moins riche, celle de Sup-Demi-Treillis :

Proposition I.2
==========

L'ensemble $\mathcal{P}(\Sigma)$ muni de l'opérateur U :

$$U : \mathcal{P}(\Sigma) \times \mathcal{P}(\Sigma) \longrightarrow \mathcal{P}(\Sigma)$$
$$(A,B) \longmapsto A \cup B \text{ défini par [I.20]}$$

constitue un Sup-Demi-Treillis, c'est-à-dire, il vérifie les conditions suivantes, quelques soient A,B, C $\in \mathcal{P}(\Sigma)$:

i. $A \cup A = A$ (idempotence)
ii. $A \cup B = B \cup A$ (commutativité)
iii. $A \cup (B \cup C) = (A \cup B) \cup C$ (associativité)

Démonstration

C'est une conséquence directe des propriétés élémentaires de l'opérateur union. (F.D.)

Comme on le voit, les propositions I.1 sur $\mathbb{P}(\Omega)$, et I.2 sur $\mathcal{P}(\Sigma)$ sont tout à fait analogues. Quelle est donc la relation entre $\mathbb{P}(\Omega)$ et $\mathcal{P}(\Sigma)$?

Proposition I.3
==========

Soient A, B $\in \mathcal{P}(\Sigma)$; soient P_A, P_B et $P_{A \cup B}$ les partitions de Ω associées aux variables A, B, et A$\cup$B respectivement. Alors :

i. $A \subseteq B \Rightarrow P_A \geqslant P_B$

donc la variable vectorielle B est plus fine que A.

ii. $P_{AUB} = P_A \wedge P_B$

Démonstration

On définit les quatre variables vectorielles suivantes :

X : dont les composantes sont les variables de l'ensemble A\B

Y : dont les composantes sont les variables de l'ensemble A$\cap$B

Z : dont les composantes sont les variables de l'ensemble B\A

W : dont les composantes sont les variables de l'ensemble A$\cup$B

$$X : \Omega \longrightarrow M_{A \setminus B} \qquad \omega \longmapsto X(\omega)$$

$$Y : \Omega \longrightarrow M_{A \cap B} \qquad \omega \longmapsto Y(\omega)$$

$$Z : \Omega \longrightarrow M_{B \setminus A} \qquad \omega \longmapsto Z(\omega)$$

$$W : \Omega \longrightarrow M_{A \cup B} \qquad \omega \longmapsto W(\omega)$$

Soit $\alpha \in M_A$. On appelle $[\alpha]_{A\setminus B}$ et $[\alpha]_{A\cap B}$ la projection du vecteur α sur les espaces $M_{A\setminus B}$ et $M_{A\cap B}$. Pour $\beta \in M_B$, on définit $[\beta]_{B\setminus A}$ et $[\beta]_{A\cap B}$ d'une façon similaire.

.i. $A \subseteq B \Rightarrow P_B \leqslant P_A$:

$P_B := \{\{\omega : B(\omega)=\beta\} : \beta \in M_B\} \setminus \{\emptyset\}$. Soit $\{\omega : B(\omega)=\beta\} \in P_B$, pour un $\beta \in M_B$:

$\{\omega : B(\omega)=\beta\} =$

$= \{\omega : Y(\omega) = [\beta]_{A\cap B} \text{ et } Z(\omega)=[\beta]_{B\setminus A}\}$

$\subseteq \{\omega : Y(\omega) = [\beta]_{A\cap B}\}$

$= \{\omega : A(\omega) = [\beta]_A\} \in P_A$, puisque $A \subseteq B \Rightarrow A\cap B = A$, donc $Y \equiv A$

.ii. $P_{A\cup B} = P_A \wedge P_B$:

$P_A := \{\{\omega : A(\omega)=\alpha\} : \alpha \in M_A\} \setminus \{\emptyset\} =$

$= \{\{\omega : X(\omega)=[\alpha]_{A\setminus B} \text{ et } Y(\omega)=[\alpha]_{A\cap B}\} : \alpha \in M_A\} \setminus \{\emptyset\} =$

$= \{\{\omega : X(\omega)=\varepsilon \text{ et } Y(\omega) = \delta\} : \varepsilon \in M_{A\setminus B} \text{ et } \delta \in M_{A\cap B}\} \setminus \{\emptyset\}.$

$P_B := \{\{\omega : B(\omega)=\beta\} : \beta \in M_B\} \setminus \{\emptyset\} =$

$= \{\{\omega : Y(\omega)=[\beta]_{A\cap B} \text{ et } Z(\omega)=[\beta]_{B\setminus A}\} : \beta \in M_B\} \setminus \{\emptyset\} =$

$= \{\{\omega : Y(\omega)=\delta' \text{ et } Z(\omega)=\gamma\} : \delta' \in M_{A\cap B} \text{ et } \gamma \in M_{B\setminus A}\} \setminus \{\emptyset\}.$

$P_A \wedge P_B = \{F \cap G : F \in P_A \text{ et } G \in P_B\} \setminus \{\emptyset\}$ d'après [I.5]

$$= \left\{ \{\omega : X(\omega)=\varepsilon \text{ et } Y(\omega)=\delta\} \cap \{\omega : Y(\omega)=\delta' \text{ et } Z(\omega)=\gamma\} : \begin{array}{l} \varepsilon \in M_{A\setminus B}, \delta \in M_{A\cap B}, \\ \delta' \in M_{A\cap B}, \gamma \in M_{B\setminus A} \end{array} \right\} \setminus \{\emptyset\}$$

$$= \{\{\omega : X(\omega)=\varepsilon \text{ et } Y(\omega)=\delta \text{ et } Y(\omega)=\delta' \text{ et } Z(\omega)=\gamma\} : \varepsilon \in M_{A\setminus B},\ \delta, \delta' \in M_{A\cap B}, \gamma \in M_{B\setminus A}\} \setminus \{\emptyset\}$$

et puisque il n'ya pas de ω pour lesquels $Y(\omega)=\delta$ et $Y(\omega)=\delta'$ avec $\delta \neq \delta'$,

on a :

$$P_A \wedge P_B = \{\{\omega : X(\omega)=\varepsilon \text{ et } Y(\omega)=\delta \text{ et } Z(\omega)=\gamma\} : \varepsilon \in M_{A\setminus B}, \delta \in M_{A\cap B}, \varepsilon \in M_{B\setminus A}\} \setminus \{\emptyset\}$$

$$= \{\{\omega : W(\omega)=(\varepsilon,\delta,\gamma)\} : (\varepsilon,\delta,\gamma) \in M_{A\cup B}\} \setminus \{\emptyset\}$$

$$= \{\{\omega : W(\omega)=\Theta\} : \Theta \in M_{A\cup B}\} \setminus \{\emptyset\} =: P_{A\cup B}$$

(F.D.)

Cette proposition termine notre discussion sur les Systèmes-Données. Nous allons étudier au paragraphe suivant l'estimation de la distribution de probabilité des variables $S \in \mathcal{P}(\Sigma)$, pour pouvoir introduire au chapitre II une classe de fonctions $H : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R}$ (ou plus généralement $H : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$) appellées les semi-valuations.

4.3 - SYSTEME-GENERATEUR (Niveau 2)

Dans notre démarche "vers le niveau 3", on passe par certains concepts qui sont une forme très générale de Système-Générateur: les distributions de probabilité conjointes. En effet, connaissant les

lois de distribution des probabilités des variables, on peut obtenir un tableau de données par simulation.

Il faut toutefois remarquer [RICH 75, chap.III] que l'utilisation des concepts probabilistes n'implique pas une nature stochastique du système étudié. Nous les utilisons non pas dans un but de modélisation ou d'identification, mais tout simplement pour savoir si certaines variables sont ou non indépendantes de certaines autres.

4.3.1. - Estimation des distributions de probabilité par les fréquences relatives

Dans sa forme la plus générale [PARZ 60], une *Mesure de Probabilité* sur un ensemble fini Ω est une fonction

$$P_r : \mathcal{P}(\Omega) \longrightarrow \mathbb{R} \qquad [I.22]$$

$$A \longmapsto P_r(A)$$

qui vérifie les propriétés :

i. $\forall A \in \mathcal{P}(\Omega) : P_r(A) \geq 0$

ii. Si $A_1, A_2, \ldots, A_K$ sont K ensembles disjoints ($\forall i \neq j : A_i \cap A_j = \emptyset$) :

$$P_r\left(\bigcup_{i=1}^{K} A_i\right) = \sum_{i=1}^{K} P_r(A_i)$$

iii. $P_r(\Omega) = 1$

Ainsi, si $X : \Omega \longrightarrow M_X$ est une variable quelconque définie sur Ω, on définit pour $\alpha \in M_X$

$$Pr(X=\alpha) := Pr(X^{-1}(\alpha)) := Pr(\{\omega : X(\omega)=\alpha\}) \qquad [I.23]$$

Toujours pour Ω fini , la définition [I.22] de probabilité revient à donner un *poids* p_j à chaque $\omega_j \varepsilon \Omega$ [CAIL 76], avec $\forall_j : p_j \geq 0$, et $\sum_j p_j = 1$; on définit alors $Pr(A) = \sum_{j:\omega_j \varepsilon A} p_j$. Mais en général, les observations ω_j ont toutes la même importance et sont de ce fait affectées de poids égaux : $\forall_j : p_j := 1/|\Omega|$. Ainsi :

$$Pr(A) := \sum_{j:\omega_j \varepsilon A} p_j = \sum_{j:\omega_j \varepsilon A} 1/|\Omega| = |A|/|\Omega|.$$

C'est cette affectation de poids égaux à tous les $\omega_j \varepsilon \Omega$ qui est à la base de l'estimation des probabilités par les fréquences relatives. Voyons ceci séparément pour les cas statiques et dynamiques.

. Systèmes Statiques

Pour tous les $S^E \varepsilon \mathcal{P}^*(\Sigma^E)$, et pour tous les $\alpha \varepsilon M_{S^E}$, on calcule

$$Pr(S^E=\alpha) := \frac{|\{\omega : S^E(\omega)=\alpha\}|}{|\Omega^E|} \qquad [I.24]$$

$$= \frac{\text{Nombre de } \omega \text{ pour lesquels } S^E(\omega)=\alpha \text{ dans le TDE}}{L}$$

. Systèmes Dynamiques

Pour tous les $S^D \varepsilon \mathcal{P}^*(\Sigma^D)$ et pour tous les $\alpha \varepsilon M_{S^D}$, on calcule

$$P_r(S^D{=}\alpha) := \frac{|\{\omega : S^D(\omega){=}\alpha\}|}{|\Omega^D|} \qquad [I.25]$$

$$= \frac{\text{Nombre de } \omega \text{ pour lesquels } S^D(\omega){=}\alpha \text{ dans le TDD}}{L - p + 1}$$

Pour avoir une estimation correcte des probabilités par les fréquences relatives, il faut en principe que certaines conditions soient satisfaites.

Aussi bien pour le cas statique que pour le cas dynamique, il faudrait que Ω soit "assez grand". Faisant référence à plusieurs auteurs, [CONA 72] donne la borne inférieure

$$|\Omega| \geq \prod_{i=1}^{N} (|M_i| - 1) \qquad [I.26]$$

De son côté, [CRON 63] suggère que cinq fois cette borne est largement suffisant. En réalité, tant que l'on ne suppose pas une distribution de probabilité (normale,Poisson, etc ...), il est impossible de faire une estimation certaine du nombre d'échantillons nécessaires. Mais puisque nous ne voulons pas nous limiter à une distribution particulière, nous supposons que le nombre d'échantillons est choisi toujours aussi grand que possible, et selon l'expérience du praticien, les possibilités d'observation du système, la puissance de calcul de l'ordinateur utilisé, etc .

Pour le cas dynamique il faudrait aussi, en toute rigueur, que le système soit stationnaire, érgodique, et que la période Δt d'échantillonnage soit fixée par le théorème de Shannon [LARM 75; ASH 65, chap.6].

Comme le dit [CONA 76], les strictes hypothèses mathématiques que la théorie des probabilités et des processus stochastiques demande, sont souvent non satisfaites et très souvent difficilement vérifiables dans le monde réel. Le choix est alors de ne rien faire (ou très peu)

au nom de la rigueur mathématique,ou bien, d'utiliser quand même ces théories formelles avec prudence, tout en sachant que les résultats obtenus ne seront jamais définitifs, et qu'il n'y a que la pratique et la confrontation avec la réalité qui pourront les confirmer.

Enfin, le problème d'estimation des distributions de probabilité a été largement étudié [CHOW 68, KU 69, WAGN 75, YOUN 78, MORG 81, la bibliographie cité par DUBU 80], même pour le cas spécifique où celles-ci sont utilisées pour calculer des entropies [MILL 55, CRON 63, PFAF 71, YVIN 78].

4.3.2. - Sur le calcul des fréquences relatives dans un tableau de Données

Le calcul des fréquences relatives pour estimer les probabilités multidimensionnelles est un problème simple, mais dont le nombre d'opérations nécessaires (donc le temps de calcul) croit très vite avec le nombre N de variables; de son côté, la quantité de mémoire requise augmente aussi très vite avec le nombre de modalités de chaque variable. C'est le cas en particulier des systèmes dynamiques où le nombre de modalités d'une variable X_i^D croit exponentiellement avec l'ordre p du système.

Ce calcul des fréquences relatives constitue le point le plus critique de l'Analyse Structurale basée sur la théorie de l'Information; par contre, lorsque on utilise d'autres indices non issus de l'entropie les besoins de calcul seront beaucoup moins importants [STAR 81]. Alors le praticien sera parfois amené à réduire de nombre de variables considérées et/ou le nombre de modalités qu'elles prennent. Bien sur, cela ne se fera pas sans perte d'information, mais on pourrait diminuer cette perte ainsi :

- on peut regrouper plusieurs variables primaires que l'on croit très liées en une seule variable primaire vectorielle. On prendrait comme ensemble de modalités le produit des ensembles des modalités des variables regroupées. Si les variables sont fortement liées entre elles, on les aurait de toutes façons retrouvées réunies dans un même sous-système, lors d'une décomposition en sous-systèmes faiblement couplés.

- on peut regrouper en une seule plusieurs des modalités d'une variable [DUFO 76]. Cela correspond à une diminution de la capacité d'observation de la variable en question. Dans la mesure du possible, il faut toujours chercher à rendre plus ou moins équiprobables toutes les modalités de la variable [CONA 76]. On verra que c'est dans le cas équiprobable que l'on tire le plus d'information du système.

Une fois connues les distributions de probabilité des variables, on peut calculer facilement leur entropie, qui représente la quantité d'information qu'une variable contient. Le chapitre suivant, consacré entièrement à la définition et aux propriétés de l'entropie et des indices qui en découlent, peut être considéré comme relevant du niveau épistémologique 2, celui du système générateurs, puisque aucune autre hypothèse ou connaissance sur le système ne sera nécessaire.

CHAPITRE II

THÉORIE DE L'INFORMATION ET ANALYSE DES SYSTÈMES.

GÉNÉRALISATION AUX SEMI-VALUATIONES ET SEMI-MODULATIONS

I - INTRODUCTION

La Théorie de l'Information (TI) est née en 1948 avec les travaux de Claude E.SHANNON [1948], et de Norbert WIENER. C'était une réponse puissante à un problème très concrêt: comment mesurer l'information et en assurer la transmission entre deux points - un émetteur et un récepteur - à travers un canal soumis au bruit (pour une présentation plus moderne et structurée de la TI classique voir [ASH 65]). Cette théorie eut un développement rapide et fécond [SLEP 74], et aujourd'hui on pourrait distinguer trois axes principaux de recherche :

- Communication : étudie les problèmes de transmission, codage, radar, bruits, codes secrets, ... (pour un bon résumé et bibliographie voir [WYNE 81]).

- Mathématique : étude abstraite de la notion d'information, et des généralisations de plus en plus sophistiquées aux différents types d'espaces mathématiques. On remarque en particulier les travaux de J.KAMPE DE FERRIET et de B.FORTE (pour un traitement complet et une large bibliographie voir [GUIA 77]; voir aussi [LOSF 74; CNRS 77; SALL 79] pour une bibliographie française).

- Analyse de systèmes : étudie les échanges d'information entre les différentes parties d'un système. C'est dans cette optique que nous présentons quelques résultats dans ce chapitre.

La bibliographie dans les deux premiers aspects mentionnés est très abondante, surtout pour l'aspect communication. Par contre dans le domaine de l'Analyse Informationnelle des systèmes, on pourrait dire qu'elle est plutôt rare, en particulier dans les dernières années. C'est pourquoi, en plus des références citées dans le texte, il nous semble intéressant de mentionner ici une série d'articles

présentant la Théorie de l'Information comme une forme de réflexion et une logique qui permet de regarder les systèmes d'un point de vue très enrichissant. A ce propos W.R.ASHBY disait " ... il existe un besoin très urgent de méthodes qui nous donnent ce que nous pouvons *vraiment* prendre, ce dont nous avons *réellement* besoin et non pas ce que nous croyons en vouloir. Trouver et développer de telles méthodes c'est le véritable problème de la cybernétique d'aujourd'hui. Une de ces méthodes est constituée par l'étude des systèmes sous leurs aspects informationnels ... " [ASHB 65] . R.C. CONANT ajoute à son tour "... même lorsque les calculs deviennent impossibles, l'interprétation informelle des équations informationnelles fournit des renseignements précieux sur le comportement des systèmes ... " [CONA 76].

Conçue originalement pour l'étude de l'information entre deux variables, message émis et message reçu, W.J. MCGILL et W.R. GARNER font la généralisation de la TI à N variables [MCGI 54; GARN 56]; W.R.ASHBY étudie des lois de transmission et d'évolution de l'information dans les systèmes complexes [ASHB 56,65,69]; W.R.ASHBY, R.C. CONANT, H.L.WEIDEMANN, B.PORTER, W.J.M. KICKERTet al, étudient les relations très importantes qui existent entre, d'un côté la transmission et la "disponibilité" de l'information, et de l'autre la capacité de régulation et de contrôle d'un système [ASHB 58; CONA 69; WEID 69; PORT 76; KICK 78] ; R.C. CONANT et W.J.M. KICKERT font, à l'aide de la TI, un rapprochement entre la régulation et la modélisation d'un système [CONA 70; KICK 78] ; D.D. KOUVATSOS étudie l'aide qui peut fournir la TI à la conception d'un système multicritère [KOUV 76] ; R.C. CONANT définit les différents traitements d'information dans un système général : il distingue les tâches de sélection de l'information significative, de blocage de celle qui ne l'est pas, de transmission interne de l'information de coordination, et de transmission vers l'extérieur des informations créées ou relayées par le système. Il montre que ces tâches sont concurrentes, c'est-à-dire qu'une augmentation de la performance d'une d'entre elles entraîne la diminution proportionnelle des autres. Cela a d'importantes conséquences sur les performances

qu'un système peut atteindre sur les objectifs voulus lorsqu'il doit évoluer dans différents types d'environnements [CONA 76]. Enfin, signalons que nombre de ces chercheurs critiquent ouvertement la communauté d'automaticiens, qu'ils appellent "classiques", pour avoir oublié ou sous-estimé l'approche informationnelle des systèmes [WEID 69, ASHB 73, PORT 76, KICK 78, ...].

Il y a plusieurs raisons qui font que la TI se prête si bien à l'analyse de systèmes de tous types, tout particulièrement lorsque l'on possède peu de connaissances à priori :

- les seules hypothèses requises sont :

  . Nombre d'échantillons suffisant pour une estimation correcte des probabilités (ce qui impose un temps d'observation suffisamment long pour les systèmes dynamiques).

  . pour les systèmes dynamiques : stationnarité et ergodicité, ainsi qu'une période d'échantillonnage appropriée [LARM 75].

- la TI s'applique à tout type de variable (numériques, semi-numériques, qualitatives, logiques, ...) puisque elle est indépendante des valeurs mêmes des variables, et prend en compte seulement la partition de l'espace d'échantillonnage Ω à laquelle chaque variable donne lieu.

- les indices de la TI sont très sensibles à n'importe quel type de relation entre les variables (linéaire, non linéaire, logique ,...) ce qui n'est pas le cas pour tous les indices issus de la statistique [BLAC 68].

- les équations de la TI ont toutes une interprétation intuitive très naturelle, que nous essayerons de mettre en valeur. D'ailleurs, dans la plupart des cas, il est possible de les interpréter graphiquement.

- Enfin, on serait même tenté de dire qu'il est possible de "prendre

quelques libertés" avec les hypothèses d'ergodicité, stationnarité et nombre d'échantillons [CONA 76]. On peut trouver quelques exemples d'application des méthodes informationnelles à des systèmes qui ne vérifient nullement les hypothèses mentionnées, et pour lesquels les résultats obtenus ont été très encourageants,d'autant plus que d'autres méthodes classiques n'avaient presque rien donné [CONA 73; STEI 74; RINK 75].

Néanmoins, ce chapitre ne se limite pas à l'étude de la théorie de l'Information. En faisant appel au formalisme des semi-valuations proposé par Zohir ABID [1979], nous essayonsde faire la présentation des résultats de façon à faciliter sa généralisation immédiate à des domaines autres que la Théorie de l'Information. Nous utilisons aussi un autre concept, celui de semi-modulation,qui nous permet de caractériser quelques propriétés peu connues de l'entropie.

Cette approche en termes de semi-valuations et semi-modulations requiert un soin particulier au niveau des hypothèses utilisées dans chaque démonstration. Nous distinguons ainsi 4 groupes d'hypothèses :

- SV : semi-valuation
- SVM : semi-valuation monotone
- SMP : semi-modulation positive
- SMPM : semi-modulation positive monotone

avec lesquelles nous "étiquetons" chaque proposition.

Cependant, ce soin permanent des hypothèses utilisées, ainsi que notre souci de donner toujours une interprétation intuitive à chacun des résultats, ne va pas sans alourdir un peu la rédaction du chapitre. Nous espérons que cela sera justifié par les généralisations qu'on pourra faire ultérieurement.

2 - DÉFINITION ET PROPRIÉTÉS DE L'ENTROPIE

2.1 - UNE MESURE INTUITIVE DE L'INCERTITUDE ET/OU DE L'INFORMATION

Soit X une variable aléatoire discrête qui prend les modalités a_1, a_2, a_3,..., a_n, avec les probabilités p_1, p_2,... p_n respectivement. Quelle est notre *incertitude à priori* vis à vis de X ?. Convenons de mesurer cette incertitude par le nombre minimal moyen de questions binaires (de réponse OUI ou NON) qu'il nous faudrait poser pour déterminer la valeur de X. Voyons tout d'abord le cas où toutes les modalités sont équiprobables, c'est-à-dire, $p_i = 1/n$ pour $i=1,2,\ldots,n$.

Lorsque la réponse à une question binaire sera OUI on aura réduit le nombre d'alternatives de n à K, pour un certain K entre 1 et n. Si par contre la réponse est NON, le nombre d'alternatives se réduira de n à (n-K). Ainsi, la probabilité d'avoir un OUI est K/n, et celle d'avoir un NON est (n-K)/n. Alors, le nombre moyen d'alternatives qu'on aura une fois la réponse connue sera :

$$K.K/n + (n-K) \cdot (n-K)/n = K^2/n + (n-K)^2/n.$$

Si l'on minimise cette expression par rapport à K, on trouve que le minimum est atteint pour K = n/2. Ainsi, ce qu'on peut faire de mieux, en moyenne, c'est poser des questions binaires qui réduisent le nombre d'alternatives de moitié. Appelons-les questions optimales.

Si l'on pose successivement h questions optimales, on réduit le nombre d'alternatives à $n/2^h$.

On aura trouvé la valeur de X lorsqu'il ne reste qu'une seule alternative, c'est-à-dire lorsque $n/2^h = 1$, donc pour $h = \log_2 n$. Ainsi, déterminer une parmi n modalités équiprobables requiert en moyenne un minimum de $\log_2 n$ questions optimales.

Supposons maintenant que les modalités $a_1, a_2, \ldots, a_n$ se produisent avec les probabilités $p_1, p_2, \ldots, p_n$ respectivement, où $\forall i : p_i \geq 0$ et $\sum_{i=1}^{n} p_i = 1$. Quelle est la situation dans ce cas ? . Le nombre minimal de questions nécessaires pour déterminer une modalité de probabilité p_i doit être le même que s'il s'agissait d'en déterminer une parmi $1/p_i$ modalités équiprobables, c'est-à-dire $\log_2 1/p_i$. Celà pour la i-ème modalité, qui ne se produit que avec une probabilité p_i. En général donc, le nombre minimal moyen de questions optimales nécessaires est la somme des nombres de questions requises dans chaque cas, pondérés par leurs probabilités respectives, c'est-à-dire

$$\text{Incertitude } (p_1,p_2,\ldots,p_n) := \sum_{i=1}^{n} p_i \cdot \log_2 1/p_i = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Regardons maintenant la situation d'un autre point de vue. Supposons que nous pouvons observer directement la variable X et la modalité qu'elle prend lors de ses réalisations. Combien d'*information à postériori* nous fournit la variable X ?. On dira que l'Information à postériori est égale au nombre minimal de questions qu'on aurait dû poser pour déterminer X, ou ce qui revient au même, à l'incertitude qu'on avait à priori vis à vis de X. C'est-à-dire

$$\text{Information } (p_1,p_2,\ldots,p_n) := -\sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Ainsi, l'Incertitude et l'Information ne sont que deux interprétations (l'une à priori, l'autre à postériori) d'un même concept qu'on appellera l'Entropie (pour une déduction formelle de l'entropie voir [ASH 65, chap.I]).

2.2 - DEFINITION FORMELLE DE L'ENTROPIE

Du fait que chaque variable définie sur Ω donne lieu à une partition de Ω, l'Entropie peut être définie de deux manières équivalentes :

. comme une fonction qui s'applique aux variables définies sur Ω, (en particulier aux variables vectorielles de $\mathcal{P}(\Sigma)$)

. comme une fonction qui s'applique aux partitions de Ω.

a) Entropie d'une variable

Soit $X : \Omega \longrightarrow M_X$ une variable quelconque définie sur Ω. On définit l'Entropie de X :

$$H(X) := - \sum_{\alpha \in M_X} Pr\ (X=\alpha).\log_2 Pr\ (X=\alpha) \text{ bits} \qquad [II.1]$$

où $Pr\ (X=\alpha) := Pr\ (X^{-1}(\alpha)) = Pr\ (\{\omega : X(\omega) = \alpha\})$.

En particulier pour l'ensemble $\mathcal{P}(\Sigma)$ des variables vectorielles dont les composantes sont des éléments de Σ, [II.1] donne lieu à une fonction :

$$H : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R} \qquad [II.2]$$

$$S \longmapsto H(S) := - \sum_{\alpha \in M_S} Pr\ (S=\alpha).\log_2 Pr\ (S=\alpha)$$

avec $H(\emptyset) := 0$.

(par définition, $o \cdot \log_2 o := o$. Cela provient du fait que $\lim_{p \to 0^+} p.\log p = 0$).

L'entropie est mesuré en *bits*. Un bit correspond à l'entropie d'une variable qui prendrait 2 modalités, chacune avec la probabilité 1/2.

b) Entropie d'une partition de Ω

On peut remarquer que l'entropie d'une variable X ne dépend pas explicitement des modalités qu'elle prend, mais seulement des probabilités avec lesquelles cela est fait; autrement dit, du nombre de $\omega \in \Omega$ pour lesquels $X(\omega) = \alpha$, et ceci pour chaque $\alpha \in M_X$. Celà est justement la partition P_X de Ω.

Plus généralement donc, on peut considérer Ω muni d'une mesure de probabilité Pr (cf.chap.I.9). L'Entropie est définie comme la fonction

$$H : \mathbb{P}(\Omega) \longrightarrow \mathbb{R} \qquad [II.3]$$

$$R = \{R_1, R_2, \ldots, R_s\} \longmapsto H(R) := -\sum_{i=1}^{s} Pr(R_i).\log_2 Pr(R_i).$$

Que l'on regarde H comme une fonction qui s'applique à une variable X ou à la partition associée P_X est tout à fait équivalent. En effet, soit $X : \Omega \longrightarrow M_X$ une variable quelconque, et sa partition $P_X := \{\{\omega : X(\omega) = \alpha\} : \alpha \in M_X\} \setminus \{\emptyset\}$. Alors

$$H(P_X) := -\sum_{\alpha \in M_X} Pr(\{\omega : X(\omega) = \alpha\}).\log_2 Pr(\{\omega : X(\omega) = \alpha\}) = -\sum_{\alpha \in M_X} Pr(X = \alpha).\log_2 Pr(X = \alpha) =: H(X)$$

REMARQUE : Au chapitre I, §4.2.4, on a défini deux variables équivalentes comme celles qui donnent lieu à la même partition de Ω. On constate immédiatement que deux variables équivalentes ont la même entropie.

Quelques conventions de notation :

- Si X, Y, Z sont des variables, il est indifférent d'écrire

  H(X) ou H({X}),

  H(X,Y,Z) ou H({X,Y,Z}) ou H({X} u {Y} u {Z}), ...

- Si A est un ensemble de variables et X est une variable, on écrira H (A,X) au lieu de H (A u {X}).

- Désormais, "log" est sous-entendu logarithme en base 2. Pour les logarithmes naturels on utilise "ln".

2.3 - QUELQUES PROPRIETES DE L'ENTROPIE

Proposition II.1

Quelle que soit la variable X, $H(X) \geqslant 0$. De plus, $H(X) = 0 \Leftrightarrow$ X est une fonction constante.

Interprétation : l'information qu'une variable fournit est toujours non négative. Les seules variables qui ne fournissent aucune information sont les variables constantes (elles prennent toujours la même modalité !).

Démonstration :

La première affirmation découle directement du fait que $\forall\, p \in [0,1]$: $-p.\log p \geqslant 0$.

Pour la seconde affirmation, $H(X)=0$ **ssi** chacun des termes de la somme $-\sum_{\alpha \in M_X} P_r\ (X=\alpha).\log P_r\ (X=\alpha)$ est nul, ce qui n'arrive que pour $P_r\ (X=\alpha) = 0$ ou $P_r\ (X=\alpha) = 1$, c'est-à-dire lorsque toutes les probabilités sont nulles, sauf une qui est égale à 1; cela veut dire que X prend toujours la même modalité (F.D.)

Lemme II.1 . Soient $p_1, p_2, \ldots, p_n$ et $q_1, q_2, \ldots, q_n$ deux suites de n nombres telles que

* $p_i \geq 0$, $q_i \geq 0$ pour $i = 1,2,\ldots,n$, et

* $q_i = 0 \Rightarrow p_i = 0$, et

* $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1.$

alors $-\sum_{i=1}^{n} p_i \cdot \log p_i \leq -\sum_{i=1}^{n} p_i.\log q_i$, avec égalité ssi $\forall i : p_i = q_i$

Démonstration :
==========

Soit $I = \{ i : p_i > 0 \}$. Alors, $\forall\ i \in I : q_i > 0$.

On sait par ailleurs que $\ln x \leq x-1$, quelque soit $x > 0$, avec égalité seulement pour $x=1$ (Fig.II.1).

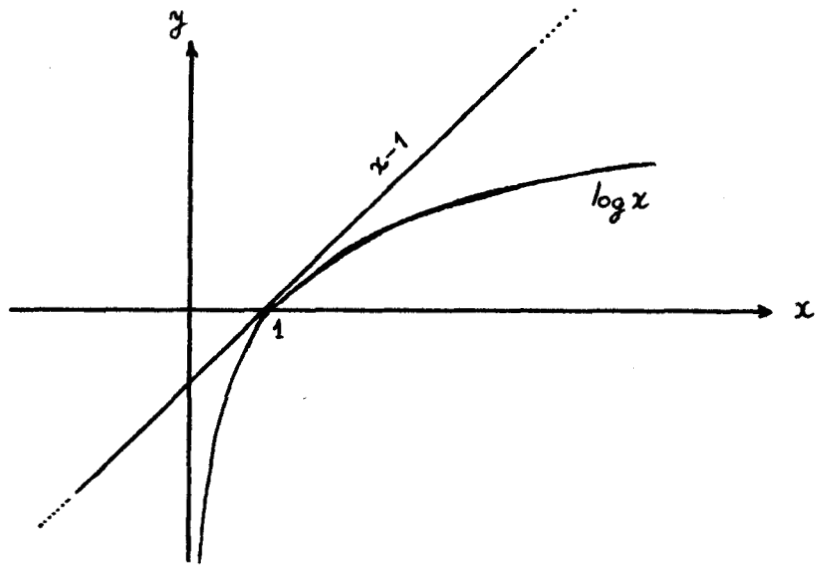


Figure II.1 : $\log x \leqslant x-1$

Alors, quelque soit $i \in I$:

$$\ln q_i/p_i \leqslant q_i/p_i - 1 \text{ avec égalité ssi } p_i = q_i$$

$$\Rightarrow p_i \cdot \ln q_i/p_i \leqslant q_i - p_i$$

$$\Rightarrow \sum_{i\in I} p_i \cdot \ln q_i/p_i \leqslant \sum_{i\in I} q_i - \sum_{i\in I} p_i = \sum_{i\in I} q_i - 1 \leqslant 0$$

$$\Rightarrow - \sum_{i\in I} p_i \cdot \ln p_i \leqslant - \sum_{i\in I} p_i \cdot \ln q_i$$

$$\Rightarrow - \sum_{i\in I} p_i \cdot \frac{\ln p_i}{\ln 2} = - \frac{1}{\ln 2} \sum_{i\in I} p_i \cdot \ln p_i \leqslant - \frac{1}{\ln 2} \sum_{i\in I} p_i \cdot \ln q_i = - \sum_{i\in I} p_i \cdot \frac{\ln q_i}{\ln 2}$$

$$\Rightarrow - \sum_{i\in I} p_i \cdot \log p_i \leqslant - \sum_{i\in I} p_i \cdot \log q_i$$

D'autre part, $- \sum_{i \notin I} p_i . \log p_i = - \sum_{i \notin I} p_i . \log q_i$

trivialement, puisque aussi bien le membre de gauche que le membre de droite sont égaux à 0. Alors

$$-\sum_{i=1}^{n} p_i . \log p_i = -\sum_{i \in I} p_i . \log p_i - \sum_{i \notin I} p_i . \log p_i \leqslant -\sum_{i \in I} p_i . \log q_i -$$

$$-\sum_{i \notin I} p_i . \log q_i = - \sum_{i=1}^{n} p_i . \log q_i ,$$

avec égalité seulement si $p_i = q_i$ (F.D.)

Proposition II.2

Soit X une variable qui prend les modalités $a_1, a_2, \ldots, a_n$, avec les probabilités $p_1, p_2, \ldots, p_n$. Alors

$$H(X) := - \sum_{i=1}^{n} p_i . \log p_i \leqslant \log n, \text{ avec égalité ssi } \forall_i \; p_i = 1/n$$

Interprétation : l'entropie (l'information reçue, l'incertitude) d'une variable est maximale lorsque toutes ses modalités sont équiprobables.

Démonstration :

On applique le lemme II.1 avec $q_i = 1/n$:

$$H(X) = - \sum_{i=1}^{n} p_i . \log p_i \leqslant - \sum_{i=1}^{n} p_i . \log \frac{1}{n} = - \log \frac{1}{n} . \sum_{i=1}^{n} p_i = \log n.$$

(F.D.)

Proposition II.3

Soient A et B deux ensembles de variables (vus comme des variables vectorielles). Alors

$$H(AuB) \leqslant H(A) + H(B)$$

avec égalité si et seulement si A et B sont des ensembles statistiquement indépendants de variables.

Interprétation : l'information qu'apporte la réunion de plusieurs groupes de variables est inférieure ou égale à la somme des informations apportées par chacun d'eux. On peut expliquer ceci par le fait qu'un groupe de variables peut contenir un peu d'information sur les autres (par exemple, connaissant l'"âge" d'un jeune enfant on a déjà une idée sur sa "taille"). Ainsi, dans H(A) + H(B) on compte deux fois cette partie commune d'information.
Et c'est seulement lorsque ces groupes de variables sont tout à fait indépendants qu'ils n'apportent aucune information répétée (la connaissance de "l'âge" d'un jeune enfant ne nous permet pas de tirer des renseignements sur son "état de santé", par exemple).

Démonstration :

On sait que :

$$Pr(A=\alpha) = \sum_{\beta \epsilon M_B} Pr(A=\alpha, B=\beta), \text{ et } Pr(B=\beta) = \sum_{\alpha \epsilon M_A} Pr(A=\alpha, B=\beta).$$

$$\text{Alors } H(A) := -\sum_{\alpha \in M_A} Pr(A=\alpha).\log Pr(A=\alpha)$$

$$= -\sum_{\alpha \in M_A} \left(\sum_{\beta \in M_B} Pr(A=\alpha, B=\beta)\right).\log Pr(A=\alpha)$$

$$= -\sum_{(\alpha,\beta) \in M_A \times M_B} Pr(A=\alpha, B=\beta).\log Pr(A=\alpha).$$

$$H(B) := -\sum_{\beta \in M_B} Pr(B=\beta).\log Pr(B=\beta)$$

$$= -\sum_{\beta \in M_B} \left(\sum_{\alpha \in M_A} Pr(A=\alpha, B=\beta)\right).\log Pr(B=\beta)$$

$$= -\sum_{(\alpha,\beta) \in M_A \times M_B} Pr(A=\alpha, B=\beta) \cdot \log Pr(B=\beta)$$

$$\text{Donc, } H(A) + H(B) = -\sum_{(\alpha,\beta) \in M_A \times M_B} Pr(A=\alpha, B=\beta).(\log Pr(A=\alpha) + \log Pr(B=\beta))$$

$$= -\sum_{(\alpha,\beta) \in M_A \times M_B} Pr(A=\alpha, B=\beta) \cdot \log Pr(A=\alpha) \cdot Pr(B=\beta)$$

$$\text{D'autre part, } H(A \cup B) := -\sum_{(\alpha,\beta) \in M_A \times M_B} Pr(A=\alpha, B=\beta).\log Pr(A=\alpha, B=\beta).$$

On peut remarquer que $\forall (\alpha,\beta) \in M_A \times M_B : Pr(A=\alpha, B=\beta) \geq 0$, et $Pr(A=\alpha).Pr(B=\beta) \geq 0$.

Aussi, si $Pr(A=\alpha).Pr(B=\beta)=0$ alors $Pr(A=\alpha)=0$ ou $Pr(B=\beta)=0$, ce qui implique que $Pr(A=\alpha, B=\beta)=0$. Finalement,

$$\sum_{(\alpha,\beta)\in M_A \times M_B} Pr(A=\alpha, B=\beta) = 1 \ ;$$

$$\sum_{(\alpha,\beta)\in M_A \times M_B} Pr(A=\alpha)\cdot Pr(B=\beta) = \sum_{\alpha\in M_A} Pr(A=\alpha)\cdot \sum_{\beta\in M_B} Pr(B=\beta) = 1\times 1 = 1 .$$

Ainsi, nous avons les conditions du Lemme II.1, donc

$$H(AuB) = -\sum_{(\alpha,\beta)\in M_A \times M_B} Pr(A=\alpha, B=\beta).\log Pr(A=\alpha, B=\beta)$$

$$\leqslant -\sum_{(\alpha,\beta)\in M_A \times M_B} Pr(A=\alpha, B=\beta).\log Pr(A=\alpha)\cdot Pr(B=\beta) = H(A)+H(B)$$

avec égalité si et seulement si $\forall(\alpha,\beta)\in M_A \times M_B : Pr(A=\alpha, B=\beta) = Pr(A=\alpha).Pr(B=\beta)$,

c'est-à-dire, si A et B sont deux ensembles indépendants de variables.

(F.D.)

La proposition II.3 peut être étendue à K ensembles de variables :

$$H(A_1 u A_2 u \ldots . u A_K) \leqslant H(A_1) + H(A_2) + \ldots . + H(A_K) \qquad [II.4]$$

avec égalité ssi $A_1, A_2, \ldots, A_K$ sont K ensembles indépendants de variables.

En effet, $H(A_1 u A_2 u \ldots . u A_K) \leqslant H(A_1 u \ldots . u A_{K-1}) + H(A_K) \leqslant H(A_1 u A_2 u \ldots . u A_{K-2}) +$

$$+ H(A_{K-1}) + H(A_K) \leqslant \ldots \leqslant H(A_1) + H(A_2) + \ldots . + H(A_K).$$

Lemme II.2 . $\forall a, b \in R$, où $0 \leq a,b \leq 1$ et $0 \leq a+b \leq 1$, on a :

$$- (a+b).\log (a+b) \leq - a.\log a - b.\log b$$

Démonstration :
==========

Si a = b = o, l'affirmation est immédiate. Si a = o et b > o, ou si a > o et b = o, il en est de même.

Il reste seulement à montrer le cas où a > o et b > o. Pour cela il nous faut un changement de notation, afin de transformer les sommes en produits. Soit

$$\theta := a+b \qquad p := \frac{a}{a+b} \qquad q := \frac{b}{a+b}$$

Ainsi, $a = p\cdot\theta$, $b = q\cdot\theta$, $p+q=1$ et p,q,θ sont positifs.

On peut réécrire l'inégalité à montrer ainsi :

$$- \theta.\log \theta \leq - p\cdot\theta.\log p\cdot\theta - q\cdot\theta.\log q\cdot\theta$$

$$\Leftrightarrow - \theta.\log \theta \leq - p\cdot\theta.\log p - p\cdot\theta.\log \theta - q\cdot\theta.\log q - q\cdot\theta.\log \theta$$

$$\Leftrightarrow - \theta.\log \theta \leq \theta\,(-p.\log p - q.\log q) - (p+q).\theta.\log \theta$$

$$\Leftrightarrow 0 \leq \theta.(- p.\log p - q.\log q)$$

Cette dernière inégalité est vraie car d'un côté $\theta > 0$, et de l'autre le terme entre parenthèses est non négatif puisqu'il correspond à l'entropie d'une variable qui prend deux modalités avec les probabilités p et q. Cela rend vraie également l'inégalité initiale que l'on voulait démontrer.

(F.D.)

Le lemme II.2 peut être généralisé par récurrence à :

$\forall\ a_1, a_2, \ldots, a_K$ où $0 \leqslant a_i \leqslant 1$ pour $i=1,\ldots,K$, et $0 \leqslant \sum_{i=1}^{K} a_i \leqslant 1$

$$-\left(\sum_{i=1}^{K} a_i\right)\cdot\log\left(\sum_{i=1}^{K} a_i\right) \leqslant -\sum_{i=1}^{K} a_i \log a_i \qquad [II.5]$$

Proposition II.4

Soient $Q = \{ Q_1, Q_2, \ldots, Q_q \}$ et $R = \{ R_1, R_2, \ldots, R_r \}$ deux éléments de $\mathbb{P}(\Omega)$:

Si $Q \leqslant R$ alors $H(Q) \geqslant H(R)$

Démonstration :
==========

Comme nous l'avons remarqué au chapitre I, § 4.2.3., le fait que $Q \leqslant R$ implique que chaque classe de Q est contenue dans une classe de R; aussi, chaque classe de R est justement l'union des classes de Q qu'elle contient. Nous avons alors

$$H(R) := -\sum_{i=1}^{r} P_r(R_i).\log P_r(R_i)$$

$$= -\sum_{i=1}^{r} P_r\left(\bigcup_{\alpha:Q_\alpha \subseteq R_i} Q_\alpha\right).\log P_r\left(\bigcup_{\alpha:Q_\alpha \subseteq R_i} Q_\alpha\right), \text{ et du fait que les } Q_\alpha \text{ sont disjoints,}$$

$$= -\sum_{i=1}^{r} \left(\sum_{\alpha:Q_\alpha \subseteq R_i} P_r(Q_\alpha)\right).\log\left(\sum_{\alpha:Q_\alpha \subseteq R_i} P_r(Q_\alpha)\right).$$

Pour chaque i, les $Pr(Q_\alpha)$ où $Q_\alpha \subseteq R_i$ vérifient les conditions du Lemme II.2. généralisé, c'est-à-dire $Pr(Q_\alpha) \geq 0$ et $0 \leq \sum_{\alpha : Q_\alpha \subseteq R_i} Pr(Q_\alpha) = Pr(R_i) \leq 1$, donc

$$H(R) = \sum_{i=1}^{r} \left[\left(- \sum_{\alpha : Q_\alpha \subseteq R_i} Pr(Q_\alpha) \right) . \log \left(\sum_{\alpha : Q_\alpha \subseteq R_i} Pr(Q_\alpha) \right) \right]$$

$$\leq \sum_{i=1}^{r} \left[- \sum_{\alpha : Q_\alpha \subseteq R_i} Pr(Q_\alpha) . \log Pr(Q_\alpha) \right] = - \sum_{i=1}^{r} \sum_{\alpha : Q_\alpha \subseteq R_i} Pr(Q_\alpha) . \log Pr(Q_\alpha)$$

$$= - \sum_{j=1}^{q} Pr(Q_j) . \log Pr(Q_j) =: H(Q) \qquad \text{(F.D.)}$$

Proposition II.5

L'entropie vérifie les deux propriétés suivantes de monotonicité :

(1) Si $X : \Omega \longrightarrow M_X$ et $Y : \Omega \longrightarrow M_Y$ sont deux variables définies sur Ω, alors

X plus fine que $Y \Rightarrow H(X) \geq H(Y)$

(2) Si $A, B \in \mathcal{P}(\Sigma)$, alors
$A \subseteq B \Rightarrow H(A) \leq H(B)$.

Interprétation : On avait remarqué au chapitre I,§ 4.2.4., que si X était plus fine que Y, alors il existait une fonction $f:M_X \longrightarrow M_Y$ telle que $Y=f(X)$. En d'autres termes, connaissant X il est alors possible de connaître Y. Il est donc normal que X apporte au moins autant d'information que Y. Quant à l'affirmation ② , il est évident que l'information et/ou l'incertitude doit augmenter lorsque l'ensemble de variables devient plus grand.

Démonstration :

① X plus fine que Y $:\Longleftrightarrow P_X \leqslant P_Y$

Alors, d'après la proposition II.4

$H(X) = H(P_X) \geqslant H(P_Y) = H(Y)$.

② D'après la proposition I.3 :

$A \subseteq B \Rightarrow P_A \geqslant P_B$, et d'après la proposition II.4 :

$H(A) = H(P_A) \leqslant H(P_B) = H(B)$ (F.D.)

On a présenté dans ce paragraphe l'entropie comme une mesure de l'information contenue dans (ou de l'incertitude vis à vis de) un ensemble de variables. On a aussi montré un certain nombre de ses propriétés dont l'interprétation intuitive est très claire. Deux de ces propriétés vont nous intéresser spécialement :

α) $\forall A,B \in \mathcal{P}(\Sigma) : H(AuB) \leqslant H(A) + H(B)$ (sous-additivité)

β) Pour $A,B \in \mathcal{P}(\Sigma)$: $A \subseteq B \Rightarrow H(A) \leqslant H(B)$ (monotonicité).

Les propriétés α et β ne sont pas exclusives à l'entropie: il y a bien d'autres fonctions définies sur $\mathcal{P}(\Sigma)$ qui les vérifient aussi. Ce sera le sujet du paragraphe suivant.

3 - SEMI-VALUATIONS, TRANSINFORMATION ET INTERVALUATION

Nous allons introduire dans ce paragraphe le concept de *semi-valuation*, dont Zohir ABID [1979] a montré dans sa thèse l'importance pour l'Analyse Structurale, puisqu'il permet d'harmoniser sous un même formalisme des approches à première vue différentes (Analyse par l'Entropie, l'Inertie, les Indices de ressemblance, ...). Désormais donc, nous continuerons notre discussion dans le cadre des semi-valuations, mais par un souci de clarté (au lecteur de juger !), les interprétations seront toujours faites en termes de la théorie de l'Information.

De plus, et cela dans le but de faciliter ces généralisations, nous accorderons une attention très spéciale aux hypothèses requises dans la démonstration de chaque proposition. Cela nous conduit à "étiqueter" chaque proposition avec l'un des sigles

- SV : proposition valable pour toute semi-valuation
- SVM : " " " " semi-valuation Monotone
- SMP : " " " " semi-modulation Positive
- SMPM: " " " " semi-modulation Positive Monotone

Les deux derniers concepts, SMP et SMPM, ne seront introduits qu'au paragraphe 7. Si une proposition n'a pas d'étiquette, ce sera parce qu'elle concerne seulement l'entropie.

3.1 - DEFINITION DES SEMI-VALUATIONS

Une fonction

$$H : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R}$$
$$S \longmapsto H(S)$$

est appelée *Semi-Valuation* (SV) si elle vérifie

$$\forall A,B \in \mathcal{P}(\Sigma) : H(A \cup B) \leqslant H(A) + H(B) \qquad [II.6]$$

On dira qu'il s'agit d'une *Semi-Valuation Monotone* (SVM) si, de plus :

$$\forall A, B \in \mathcal{P}(\Sigma) : A \subseteq B \Rightarrow H(A) < H(B) \qquad [II.7]$$

REMARQUE : La propriété [II.6] peut en fait être étendue immédiatement par récurrence à :

$$\forall A_1,A_2,\ldots,A_K \in \mathcal{P}(\Sigma): H(A_1 \cup A_2 \cup \ldots \cup A_K) \leqslant H(A_1)+H(A_2)+\ldots+H(A_K) \qquad [II.8]$$

Proposition II.6 (SV)

Quelle que soit $A \in \mathcal{P}(\Sigma)$: $H(A) \geqslant 0$

Démonstration :

$$H(A) = H(A \cup A) \leqslant H(A) + H(A) = 2 \cdot H(A)$$

$$\Rightarrow \quad H(A) \geqslant 0 \qquad \text{(F.D.)}$$

3.2 - TRANSINFORMATION ET INTERVALUATION

3.2.1. - Cas H=Entropie : les transinformations

L'inégalité $H(AuB) \leq H(A) + H(B)$ avait été interprétée par le fait que dans $H(A) + H(B)$ on compte deux fois une partie d'information qui se trouve aussi bien en A qu'en B, et qu'on suppose donc comme ayant été transmise. Pour mesurer cette transmission d'information, on définit :

- *Transinformation Interne :*

$$\dot{I} : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R} \qquad [II.9]$$
$$S \longmapsto \dot{I}(S) := \sum_{X \in S} H(X) - H(S)$$

qui indique la quantité d'information qui est transmise *dans* un groupe de variables ,

- *Transinformation Externe :*

$$\overleftrightarrow{I} : \mathcal{P}(\mathcal{P}(\Sigma)) \longrightarrow \mathbb{R} \qquad [II.10]$$
$$R=\{R_1,R_2,\ldots,R_r\} \longmapsto \overleftrightarrow{I}(R) = \sum_{i=1}^{r} H(R_i) - H\left(\bigcup_{i=1}^{r} R_i\right)$$

qui indique la quantité d'information qui est transmise *entre* plusieurs groupes de variables.

3.2.2. - Cas H=Semi-Valuation quelconque : les Intervaluations

Lorsqu'on applique les définitions [II.9] et [II.10] à des semi-valuations H autres que l'entropie, on appelle $\dot{I}$ l'*Intervaluation Interne* et $\overleftrightarrow{I}$ l'*Intervaluation Externe* .

Conventions de notation :

L'argument de $\overleftrightarrow{I}$ est un ensemble de sous-ensembles de variables. Toutefois, si cela ne prête pas à confusion, on peut faire quelques simplifications de notation :

Si $X_i, X_j, X_{i_1}, X_{i_2}, \ldots, X_{i_K} \in \Sigma$, et $A, A_1, A_2, \ldots, A_K \in \mathcal{P}(\Sigma)$, on écrit :

$\overleftrightarrow{I}(X_i:X_j)$ au lieu de $\overleftrightarrow{I}(\{\{X_i\},\{X_j\}\})$,

$\overleftrightarrow{I}(X_{i_1}:X_{i_2}:\ldots:X_{i_K})$ au lieu de $\overleftrightarrow{I}(\{\{X_{i_1}\},\{X_{i_2}\},\ldots,\{X_{i_K}\}\})$,

$\overleftrightarrow{I}(A_1:A_2:\ldots:A_K)$ au lieu de $\overleftrightarrow{I}(\{A_1,A_2,\ldots,A_K\})$.

Par ailleurs,

$\dot{I}(\emptyset):=0$; $\overleftrightarrow{I}(A:\emptyset)=0$; $\dot{I}(\{X_i\}) = 0$; $\overleftrightarrow{I}(A) = 0$.

REMARQUE : D'après leurs définitions, il est évident que

$\overleftrightarrow{I}(A_1:A_2:\ldots:A_K)$ et $\dot{I}(\{X_{i_1},X_{i_2},\ldots,X_{i_K}\})$

ne dépendent pas de l'ordre des arguments.

4 - PROPRIÉTÉS DES INTERVALUATIONS

Une bonne partie des propriétés que nous allons présenter dans ce paragraphe ont été étudiées par W.R. ASHBY [1965,1969] mais seulement pour H=Entropie. Ce qui est peut-être plus nouveau c'est que, en suivant l'approche proposée par Z.ABID [1979], elles seront démontrées pour des conditions plus générales.

Proposition II.7 (SV)

Si $Y_1, Y_2, \ldots, Y_K$ sont K variables différentes :

$$\overleftrightarrow{I}(Y_1 : Y_2 : \ldots : Y_K) = \dot{I}(\{Y_1, Y_2, \ldots, Y_K\})$$

Proposition II.8 (SV)

① $\forall R = \{R_1, R_2, \ldots, R_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$

$$\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1 : R_2 : \ldots : R_r) \geq 0 .$$

② $\forall A \in \mathcal{P}(\Sigma) : \dot{I}(A) \geq 0$

De plus, pour H=Entropie, l'égalité à 0 dans ① se produit ssi $R_1, R_2, \ldots R_r$ sont r ensembles indépendants de variables; et dans ② ssi toutes les variables de A sont indépendantes.

Démonstration :
==========

Ces deux affirmations sont une conséquence directe de [II.8] :

$$H(A_1 u A_2 u \ldots u A_r) \leq H(A_1) + H(A_2) + \ldots + H(A_r).$$

Par ailleurs, la proposition II.3 généralisée montre que pour H=Entropie, cette inégalité devient une égalité seulement lorsque $A_1,\ldots,A_r$ sont indépendantes.

(F.D.)

Pour H=Entropie, cette propriété des transinformations de s'annuler seulement lorsque les variables sont tout à fait indépendantes est très importante. Elle implique que dans tous les autres cas, et quel que soit le type de relation entre les variables, le transinformation indiquera une valeur positive. Cela constitue un avantage certain vis à vis d'autres indices de relation entre variables. Par exemple si on considère X et Y les variables échantillonnées de sint et |sint| , respectivement, entre 0 et 2Π,

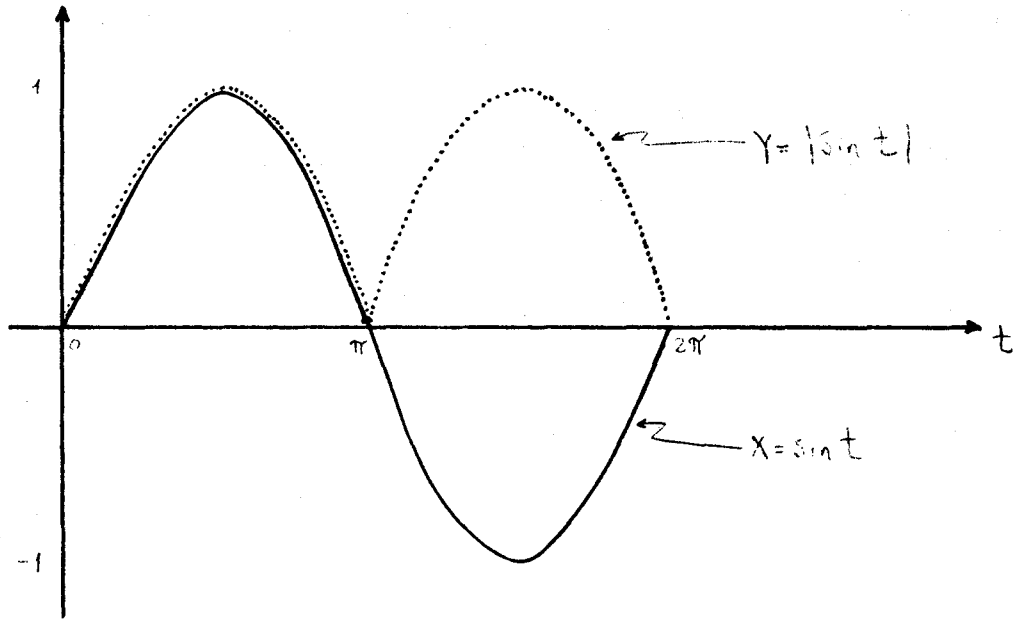


Figure II.2 - Deux fonctions dont le coefficient de corrélation est 0.

le coefficient de corrélation $\rho\ (X,Y) = 0$ (il est + 1 entre 0 et π, et -1 entre π et 2π). Mais il est évident que X et Y sont très liées. On peut montrer que

$$H(X) = H(Y) + 1 \quad ; \quad H(X,Y) = H(X)$$

Dans la première égalité, le "+1" représente justement l'information apportée par la connaissance du signe de X(qui peut être + ou - avec la même probabilité); le fait que l'information fournie par X et Y soit égale à celle fournie par X tout seul, s'explique parce que connaissant la valeur de X on connait implicitement la valeur de Y (il suffit de le rendre positif s'il ne l'est pas). Ainsi,

$$\overleftrightarrow{I}(X:Y) = H(X) + H(Y) - H(X,Y) = H(Y) > 0$$

qui exprime bien la liaison qu'on ressent intuitivement entre X et Y.

Proposition II.9 (SV)

Soit $R = \{R_1, R_2, \ldots, R_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$:

① Quel que soit $R' = \{R_{i_1}, R_{i_2}, \ldots, R_{i_p}\} \subseteq \{R_1, R_2, \ldots, R_r\} = R$ on a

$$\overleftrightarrow{I}(R') = \overleftrightarrow{I}(R_{i_1}:R_{i_2}:\ldots:R_{i_p}) \leqslant \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = \overleftrightarrow{I}(R)$$

② Si $\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = 0$, alors

$$\forall R_i, R_j \in R \quad : \quad \overleftrightarrow{I}(R_i:R_j) = 0.$$

Interprétation : en termes d'information, l'affirmation ① dit (toutes proportions gardées) que si dans une réunion de r personnes, quelques unes sortent, l'échange d'information ne peut que diminuer. L'affirmation ② dit que si dans une réunion de r personnes il n'y a pas d'échange d'information, alors, il n'y en aura pas non plus entre deux personnes quelconques du groupe. Mais ce qui est peut être curieux, c'est que la réciproque de ② n'est pas toujours vraie. Au § 9.3 on montrera un système constitué par 4 variables qui n'échangent aucune information l'une avec l'autre, et cependant l'échange global entre les 4 variables est loin d'être nul.

Démonstration :

① Soit $R'' := R \setminus R'$. Ainsi $R = R' \cup R''$ et $R' \cap R'' = \emptyset$

Soit $J_{R'} := \{i : R_i \in R'\}$ et $J_{R''} := \{i : R_i \in R''\}$

Alors $J_{R'} \cup J_{R''} = \{1,2,..,r\}$, $J_{R'} \cap J_{R''} = \emptyset$

Avec cette notation, on a :

$$\overleftrightarrow{I}(R) := \sum_{i=1}^{n} H(R_i) - H(\bigcup_{i=1}^{r} R_i)$$

$$= \sum_{i \in J_{R'}} H(R_i) + \sum_{i \in J_{R''}} H(R_i) - H(\bigcup_{i=1}^{r} R_i)$$

$$= \left[\sum_{i \in J_{R'}} H(R_i) - H(\bigcup_{i \in J_{R'}} R_i) \right] + \left[\sum_{i \in J_{R''}} H(R_i) - H(\bigcup_{i \in J_{R''}} R_i) \right] +$$

$$+ \left[H\Big(\bigcup_{i\in J_{R'}} R_i\Big) + H\Big(\bigcup_{i\in J_{R''}} R_i\Big) - H\Big(\Big(\bigcup_{i\in J_{R'}} R_i\Big) \cup \Big(\bigcup_{i\in J_{R''}} R_i\Big)\Big) \right]$$

$$=: \overleftrightarrow{I}(R') + \overleftrightarrow{I}(R'') + \overleftrightarrow{I}\Big(\bigcup_{i\in J_{R'}} R_i : \bigcup_{i\in J_{R''}} R_i\Big)$$

$$\geqslant I(R')$$

(2) Par la partie (1) qu'on vient de montrer, on sait déjà que $0=\overleftrightarrow{I}(R) \geqslant \overleftrightarrow{I}(R_i:R_j)$. Par ailleurs, la proposition II.8 dit que $\overleftrightarrow{I}$ est toujours non négative, donc $\overleftrightarrow{I}(R_i:R_j) = 0$.

(F.D.)

Proposition II.10 (SV)

Soit $A \in \mathcal{P}(\Sigma)$:

(1) quelque soit $A' \subseteq A$, on a $\dot{I}(A') \leqslant \dot{I}(A)$

(2) Si $\dot{I}(A) = 0$ alors $\forall X_i, X_j \in A$, $i \neq j$, on a

$$\dot{I}(\{X_i, X_j\}) = \overleftrightarrow{I}(X_i:X_j) = 0$$

Interprétation : similaire à la proposition II.9, celle-ci dit que l'information échangée à l'intérieur d'un ensemble de variables diminue si l'on réduit l'ensemble. Dans ce cas non plus, la réciproque de (2) n'est pas vraie pour H=Entropie.

Démonstration :
==========

① Soit $A'' = A \setminus A'$. Ainsi $A = A' \cup A''$, $A' \cap A'' = \emptyset$.

$$\dot{I}(A) := \sum_{X \in A} H(X) - H(A)$$

$$= \sum_{X \in A'} H(X) + \sum_{X \in A''} H(X) - H(A' \cup A'')$$

$$= \sum_{X \in A'} H(X) - H(A') + \sum_{X \in A''} H(X) - H(A'') + H(A') + H(A'') - H(A' \cup A'')$$

$$=: \dot{I}(A') + \dot{I}(A'') + I(A' : A'')$$

$$\geqslant \dot{I}(A')$$

② Par la partie ① qu'on vient de montrer

$0 = \dot{I}(A) \geqslant \dot{I}(\{X_i, X_j\})$, et puisque $\dot{I} \geqslant 0$ toujours,

alors $\dot{I}(\{X_i, X_j\}) = \overleftrightarrow{I}(X_i : X_j) = 0$

(F.D.)

Proposition II.11 (SVM)
---

① Soit $R = \{R_1, R_2, \ldots, R_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$. On a

$$\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1 : R_2 : \ldots : R_r) \leqslant \sum_{i=1}^{r} H(R_i) - \max_j \{H(R_j)\} \leqslant (r-1) . \max_j \{H(R_j)\}$$

en particulier $\overleftrightarrow{I}(A:B) \leq \text{Min } \{H(A),H(B)\} \leq H(AuB)$

② Soit $A \in \mathcal{P}(\Sigma)$. On a

$$\dot{I}(A) \leq \sum_{X \in A} H(X) - \max_{X \in A} \{H(X)\} \leq (|A|-1).\text{Max}_{X \in A} \{H(X)\}$$

en particulier $\overleftrightarrow{I}(X:Y) = \dot{I}(\{X,Y\}) \leq \textbf{Min } \{H(X), H(Y)\} \leq H(X,Y)$

Interprétation : pour deux arguments elle est très claire : la quantité d'information échangée entre deux variables A et B ne peut pas dépasser la quantité d'information qu'on trouve dans A ou dans B.

Démonstration :

① $H(R_1) - H(R_1 u R_2 u \ldots u R_r) \leq 0$ puisque $R_1 \subseteq R_1 u R_2 u \ldots u R_r$

$\Rightarrow H(R_1) + H(R_2) - H(R_1 u R_2 u \ldots u R_r) \leq H(R_2)$

$\Rightarrow H(R_1) + H(R_2) + H(R_3) - H(R_1 u R_2 u \ldots u R_r) \leq H(R_2) + H(R_3)$

$\vdots$

$\Rightarrow \overleftrightarrow{I}(R) := H(R_1)+H(R_2)+\ldots\ldots\ldots+H(R_r)-H(R_1 u R_2 u \ldots u R_r) \leq$

$$\leq \sum_{i=2}^{r} H(R_i) = \sum_{i=1}^{r} H(R_i) - H(R_1)$$

Puisque $\overleftrightarrow{I}(R)$ ne dépend pas de l'ordre des R_i, on peut dire que

$$\overleftrightarrow{I}(R) \leq \sum_{i=1}^{r} H(R_i) - H(R_j) \text{ quelque soit } j,\ 1 \leq j \leq r \ ,$$

alors $\overleftrightarrow{I}(R) \leqslant \sum_{i=1}^{r} H(R_i) - \max_j\{H(R_j)\} \leqslant \sum_{i=1}^{r} (\max_j\{H(R_j)\}) - \max_j\{H(R_j)\} =$

$$= (r-1).\max_j \{H(R_j)\}.$$

alors en particulier

$$\overleftrightarrow{I}(A:B) \leqslant H(A) + H(B) - \max\{H(A),H(B)\} = \min\{H(A),H(B)\} \leqslant H(AuB).$$

La démonstration de l'affirmation (2) est tout à fait analogue à celle-ci. (F.D.)

Proposition II.12 (SV)

(1) Soient $R_1,R_2,\ldots R_r \in \mathcal{P}(\Sigma)$, alors

$$\overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_K) + \overleftrightarrow{I}(\bigcup_{i=1}^{K} R_i : \bigcup_{i=K+1}^{r} R_i) + \overleftrightarrow{I}(R_{K+1}:\ldots:R_r)$$

quelque soit $K=2,3,\ldots,r-2$;

$$= \overleftrightarrow{I}(R_1:R_2uR_3u\ldots uR_r) + \overleftrightarrow{I}(R_2:R_3:\ldots:R_r)$$

$$= \overleftrightarrow{I}(R_1:R_2:\ldots:R_{r-1}) + \overleftrightarrow{I}(R_1uR_2u\ldots uR_{r-1}:R_r) .$$

(2) Soient $R_1,R_2,\ldots,R_r \in \mathcal{P}(\Sigma)$, alors

$$\overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = \overleftrightarrow{I}(R_1:R_2) + \overleftrightarrow{I}(R_1uR_2:R_3) + \overleftrightarrow{I}(R_1uR_2uR_3:R_4) + \ldots. +$$

$$+ \overleftrightarrow{I}(R_1uR_2u\ldots.uR_{r-1}:R_r)$$

③ Soient $A,A' \in \mathcal{P}(\Sigma)$, $A' \subseteq A$. Soit $A'' := A \setminus A'$. Alors

$$\dot{I}(A) = \dot{I}(A') + \overleftrightarrow{I}(A':A'') + \dot{I}(A'')$$

Démonstration :
==========

① Pour les 3 égalités, il suffit de développer le membre de droite et faire les annulations des termes de signes opposés. Par exemple pour la première :

$$\overleftarrow{I}(R_1:R_2:\ldots:R_K) + \overleftrightarrow{I}(\overset{K}{\underset{i=1}{U}} R_i : \overset{r}{\underset{i=K+1}{U}} R_i) + \overleftarrow{I}(R_{K+1}:R_{K+2}:\ldots:R_r)$$

$$:= \sum_{i=1}^{K} H(R_i) - H(\overset{K}{\underset{i=1}{U}} R_i) + H(\overset{K}{\underset{i=1}{U}} R_i) + H(\overset{r}{\underset{i=K+1}{U}} R_i) - H(\overset{r}{\underset{i=1}{U}} R_i) + \sum_{i=K+1}^{r} H(R_i) - H(\overset{r}{\underset{i=K+1}{U}} R_i)$$

$$= \sum_{i=1}^{K} H(R_i) + \sum_{i=K+1}^{r} H(R_i) - H(\overset{r}{\underset{i=1}{U}} R_i) = \sum_{i=1}^{r} H(R_i) - H(\overset{r}{\underset{i=1}{U}} R_i) =: \overleftrightarrow{I}(R_1:\ldots:R_r).$$

② Cette affirmation est obtenue par l'application répétée de la troisième égalité de ① , ainsi :

$$\overleftarrow{I}(R_1:R_2:\ldots:R_r) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_{r-1}) + \overleftrightarrow{I}(R_1 u R_2 u \ldots u R_{r-1}:R_r).$$

$$= \overleftrightarrow{I}(R_1:R_2:\ldots:R_{r-2}) + \overleftrightarrow{I}(R_1 u \ldots u R_{r-2}:R_{r-1}) + \overleftrightarrow{I}(R_1 u R_2 u \ldots u R_{r-1}:R_r).$$

$$= \overleftrightarrow{I}(R_1:R_2:\ldots:R_{r-3}) + \overleftrightarrow{I}(R_1 u R_2 u \ldots u R_{r-3}:R_{r-2}) +$$

$$+ \overleftrightarrow{I}(R_1 u R_2 u \ldots u R_{r-2}:R_{r-1}) + \overleftrightarrow{I}(R_1 u R_2 u \ldots u R_{r-1}:R_r)$$

$$= \ldots\ldots$$

$$= \overleftrightarrow{I}(R_1:R_2) + \overleftrightarrow{I}(R_1 u R_2:R_3) + \overleftrightarrow{I}(R_1 u R_2 u R_3:R_4) + \ldots +$$

$$+ \overleftrightarrow{I}(R_1 u R_2 u \ldots . u R_{r-1} : R_r).$$

(3) On développe le côté droit de l'égalité.

$$\dot{I}(A') + \overleftrightarrow{I}(A' : A'') + \dot{I}(A'')$$

$$:= \sum_{X \in A'} H(X) - H(A') + H(A') + H(A'') - H(A' \cup A'') + \sum_{x \in A''} H(X) - H(A'')$$

$$= \sum_{X \in A'} H(X) + \sum_{X \in A''} H(X) - H(A' \cup A'') = \sum_{X \in A} H(X) - H(A) =: \dot{I}(A)$$

(F.D.)

5 - PARTITIONS DE Σ ET INTERVALUATIONS

Une partition de l'ensemble de variables Σ peut être vue comme une décomposition en sous-systèmes disjoints du système étudié. Pour pouvoir disposer au chapitre III des outils nécessaires à l'étude des Systèmes-Structure partition, on établit dès maintenant quelques propriétés de $\overleftrightarrow{I}$ et $\dot{I}$ sur les partitions de Σ.

On appelle $IP(\Sigma)$ l'ensemble de toutes les partitions de Σ. Comme on l'a vu au chapitre I,§ 4.2.3., cet ensemble $IP(\Sigma)$, muni de la relation $\preccurlyeq$: "être plus fine que", constitue un treillis. Chaque élément de $IP(\Sigma)$ étant lui-même un ensemble de sous-ensembles de Σ, alors il est clair que $IP(\Sigma) \subseteq \mathcal{P}(\mathcal{P}(\Sigma))$. Nous pouvons donc appliquer $\overleftrightarrow{I}$ aux partitions de Σ.

Proposition II.13 (SV)

Quelle que soit la partition $R=\{S_1,S_2,\ldots,S_r\} \in IP(\Sigma)$:

$$\overleftrightarrow{I}(S_1:S_2:\ldots:S_r) + \sum_{i=1}^{r} \dot{I}(S_i) = \text{Constante} = \overleftrightarrow{I}(X_1:X_2:\ldots X_N)$$

Interprétation : l'information totale transmise dans un système peut être partitionnée en deux : celle transmise à l'intérieur des sous-systèmes et celle transmise entre les sous-systèmes.

Démonstration :

$$\sum_{i=1}^{r} \dot{I}(S_i) + \overleftrightarrow{I}(S_1:S_2:\ldots S_r)$$

$$:= \sum_{i=1}^{r} \left[\sum_{X \in S_i} H(X) - H(S_i) \right] + \sum_{i=1}^{r} H(S_i) - H\left(\bigcup_{i=1}^{r} S_i\right)$$

$$= \sum_{i=1}^{r} \sum_{X \in S_i} H(X) - \sum_{i=1}^{r} H(S_i) + \sum_{i=1}^{r} H(S_i) - H\left(\bigcup_{i=1}^{r} S_i\right)$$

$$= \sum_{i=1}^{N} H(X_i) - H(\Sigma) =: \overleftrightarrow{I}(X_1 : X_2 : \ldots : X_N)$$

(F.D.)

La transinformation totale dans un système Σ, $\overleftrightarrow{I}(X_1 : X_2 : \ldots : X_N)$, constitue vraisemblablement le paramètre le plus important de celui-ci [ASHB 65]. Il mesure la quantité totale de relation, la "quantité totale de loi" que l'on peut espérer extraire du système décrit par les variables $X_1, X_2, \ldots, X_N$, même avant que les détails spécifiques de ces lois aient été établis. Celà rejoint ce que disait H.A. SIMON [1962], qui affirmait que c'est seulement si les parties d'une structure sont liées, si certains aspects de la structure peuvent être inférés à partir d'autres, si la structure est redondante, qu'on pourra faire une représentation ou description simplifiée de celle-ci. Si une structure complexe ne possède pas de redondance, elle-même sera sa plus simple description.

Proposition II.14 (SV)

Soient $R=\{R_1, R_2, \ldots, R_r\}$ et $S=\{S_1, S_2, \ldots, S_s\}$ deux éléments de $\mathbb{P}(\Sigma)$

$$R \leqslant S \Rightarrow \begin{cases} \overleftrightarrow{I}(R) \geqslant \overleftrightarrow{I}(S) \\ \sum_{i=1}^{r} \dot{I}(R_i) \leqslant \sum_{i=1}^{s} \dot{I}(S_i) \end{cases}$$

Interprétation : si la décomposition en sous-systèmes devient plus fine, la transinformation entre les sous-systèmes augmente, donc, la transinformation à l'intérieur des sous-systèmes diminue.

Démonstration :

Du fait que $R \leqslant S$, chaque classe de S est l'union de certaines classes de R. Moyennant une rénumérotation, on peut obtenir une partition des classes de R décrite par :

$$R = \{ \underbrace{R_1, R_2, \ldots, R_j}, \underbrace{R_{j+1}, \ldots, R_K}, \ldots, \underbrace{R_l, R_{l+1}, \ldots, R_r} \}$$

$$S = \{ \quad S_1, \quad S_2, \quad \ldots\ldots \quad S_s \quad \}$$

en considérant maintenant chaque R_i comme une variable (vectorielle), on peut appliquer la proposition II.13, ainsi

$$\sum_{i=1}^{s} \dot{I}(\{R_\alpha : R_\alpha \subseteq S_i\}) + \overleftrightarrow{I}(S_1 : S_2 : \ldots : S_s) = \overleftrightarrow{I}(R_1 : R_2 : \ldots : R_r)$$

le premier terme du membre gauche étant positif, alors

$$\overleftrightarrow{I}(S) = \overleftrightarrow{I}(S_1 : S_2 : \ldots\ldots S_s) \leqslant \overleftrightarrow{I}(R_1 : R_2 : \ldots\ldots : R_r).$$

Cette inégalité, avec le fait que :

$$\sum_{i=1}^{s} \dot{I}(S_i) + \overleftrightarrow{I}(S_1 : S_2 : \ldots\ldots : S_s) = \sum_{i=1}^{r} \dot{I}(R_i) + \overleftrightarrow{I}(R_1 : R_2 : \ldots\ldots : R_r)$$

implique que $\sum_{i=1}^{s} \dot{I}(S_i) \geqslant \sum_{i=1}^{r} \dot{I}(R_i)$ (F.D.)

Proposition II.15 (SV)

Soit $R = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}(\Sigma)$. Soit R' la partition de Σ obtenue en regroupant en une seule les classes R_i et R_j :

$$R' = \{R_1, R_2, \ldots, R_i \cup R_j, \ldots, R_r\} = (R \setminus \{R_i, R_j\}) \cup \{R_i \cup R_j\}$$

alors $\overleftrightarrow{I}(R') = \overleftrightarrow{I}(R) - \overleftrightarrow{I}(R_i : R_j)$

Interprétation : lorsqu'on regroupe deux sous-systèmes R_i et R_j dans un seul, la transinformation externe diminue de la transmission qui existait entre R_i et R_j.

Démonstration :

$$\overleftrightarrow{I}(R) - \overleftrightarrow{I}(R_i : R_j) := \sum_{K=1}^{r} H(R_i) - H(\Sigma) - H(R_i) - H(R_j) + H(R_i \cup R_j)$$

$$= \sum_{\substack{K=1 \\ K \neq i,j}}^{r} H(R_K) + H(R_i \cup R_j) - H(\Sigma) =: I(R')$$

(F.D.)

Proposition II.16 (SV)

Soit $R = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}(\Sigma)$. Soit R' la partition de Σ obtenue en éclatant la classe R_i en deux classes non vides $R_i \setminus S, S$ où $S \subseteq R_i$, $S \neq \emptyset$, $S \neq R_i$.

$$R'=\{R_1,R_2,\ldots,R_i\setminus S,S,\ldots,R_r\} = (R\setminus\{R_i\}) \cup \{R_i\setminus S,S\}$$

alors $\overleftrightarrow{I}(R') = \overleftrightarrow{I}(R) + \overleftrightarrow{I}(R_i\setminus S:S)$

Interprétation : lorsque à partir d'un sous-système on en crée deux, la transinformation externe augmente de la transmission entre les deux sous-systèmes créés.

Démonstration :

$$\overleftrightarrow{I}(R) + \overleftrightarrow{I}(R_i\setminus S:S) := \sum_{K=1}^{r} H(R_i) - H(\Sigma) + H(R_i\setminus S) + H(S) - H(R_i)$$

$$= \sum_{\substack{K=1\\K\neq i}}^{r} H(R_K) + H(R_i\setminus S) + H(S) - H(\Sigma) =: \overleftrightarrow{I}(R')$$

(F.D.)

6 - SEMIVALUATIONS ET INTERVALUATIONS CONDITIONNELLES

6.1 - CAS H=ENTROPIE : ENTROPIE ET TRANSINFORMATION CONDITIONNELLE

La transinformation $\overleftrightarrow{I}(A:B) := H(A) + H(B) - H(AuB)$ mesure la quantité d'information que se trouve aussi bien en A qu'en B(et qu'on suppose donc transmise entre A et B). On se demande maintenant combien il nous reste à connaître d'information sur B une fois connu A. La réponse est évidente : elle est donnée par $H(B) - \overleftrightarrow{I}(A:B)$. Mais

$$H(B) - \overleftrightarrow{I}(A:B) = H(B) - H(A) - H(B) + H(AuB) = H(BuA) - H(A).$$

Cela nous conduit à définir l'*Entropie A-Conditionnelle*, quel que soit $A \in \mathcal{P}(\Sigma)$:

$$H(./A) : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R} \qquad [II.11]$$

$$B \longmapsto H(B/A) := H(BuA) - H(A)$$

Elle peut être interprétée comme la quantité d'information qui reste à connaître (l'incertitude) sur l'ensemble de variables B, une fois qu'on possède déjà l'information fournie par A.

Quelques observations immédiates ont une interprétation intuitive évidente :

$$H(B/\emptyset) := H(Bu\emptyset) - H(\emptyset) = H(B) \qquad [II.12]$$

l'entropie de B ne sachant rien se confond avec l'entropie de B;

$$H(\emptyset/A) := H(\emptyset uA) - H(A) = 0 \qquad [II.13]$$

lorsqu'il n'y a rien à connaître, la connaissance d'un A quelconque ne change rien ;

$$H(A/A) = H(AuA) - H(A) = 0 \qquad [II.14]$$

l'incertitude d'un évènement connu est nulle ;

$$\overleftrightarrow{I}(A:B) = H(A) - H(A/B) = H(B) - H(B/A) \qquad [II.15]$$

l'information qu'on trouve aussi bien en A qu'en B est égale à celle qu'il y a en A moins celle que B ne possède pas sur A (et vice versa);

$$H(AuB) = H(A) + H(B/A) = H(B) + H(B/A) \qquad [II.16]$$

l'information fournie par AuB est égale à celle fournie par A tout seul, plus celle qui reste à connaitre de B une fois connu A.

Cette dernière égalité peut être généralisée :

$$\begin{aligned} H(A_1uA_2u...uA_K) &= H(A_1uA_2u...uA_{K-1}) + H(A_K/A_1uA_2u...uA_{K-1}) \\ &= H(A_1uA_2u...uA_{K-2}) + H(A_{K-1}/A_1uA_2u....uA_{K-2}) + H(A_K/A_1uA_2u..uA_{K-1}) \\ &= .... \\ &= H(A_1) + H(A_2/A_1) + H(A_3/A_1uA_2)+ ...+ A(A_K/A_1uA_2u...uA_{K-1}) \end{aligned} \qquad [II.17]$$

Toujours pour H=Entropie, nous pouvons obtenir une expression intéressante de H(B/A). Soient

$$A:\Omega \longrightarrow M_A \quad \text{et} \quad B:\Omega \longrightarrow M_B$$

deux éléments de $\mathcal{P}(\Sigma)$, avec

$$M_A = \{\alpha_1,\alpha_2,...., \alpha_m\} \text{ et } M_B = \{\beta_1,\beta_2,..., \beta_v\}$$

Soient $P_{i.} := Pr(A=\alpha_i)$, $P_{.j} := Pr(B=\beta_j)$, $P_{ij} := Pr(A=\alpha_i, B=\beta_j)$.

on a :

$$H(B/A) := H(AuB) - H(A)$$

$$:= -\sum_{i=1}^{m} \sum_{j=1}^{v} P_{ij} \cdot \log P_{ij} + \sum_{i=1}^{m} P_{i.} \cdot \log P_{i.}$$

$$= -\sum_{i=1}^{m} \sum_{j=1}^{v} P_{ij} \cdot \log P_{ij} + \sum_{i=1}^{m} \left(\sum_{j=1}^{v} P_{ij}\right) \cdot \log P_{i.}$$

$$= -\sum_{i=1}^{m} \sum_{j=1}^{v} P_{ij} \cdot \log \frac{P_{ij}}{P_{i.}} = -\sum_{i=1}^{m} P_{i.} \left[\sum_{j=1}^{v} \frac{P_{ij}}{P_{i.}} \cdot \log \frac{P_{ij}}{P_{i.}}\right]$$

$$= \sum_{i=1}^{m} Pr(A=\alpha_i) \cdot \left[-\sum_{j=1}^{v} Pr(B=\beta_j/A=\alpha_i) \cdot \log Pr(B=\beta_j/A=\alpha_i)\right]$$

$$=: \sum_{i=1}^{m} Pr(A=\alpha_i) \cdot H(B/A=\alpha_i) \qquad [II.18]$$

A partir de cette équation, W.R.ASHBY [1965] propose une autre interprétation de l'entropie conditionnelle : H(B/A) serait la quantité moyenne d'information que fournit encore B lorsqu'on aurait rendu constante la partie du système concernée par A (et uniquement celle-ci). Ainsi par exemple, dans un système où sont en jeu des variables comme Température, Pression, Volume, ..., H(Pression/Température) serait la quantité moyenne d'information qu'on pourrait extraire encore de la variable Pression lorsqu'on aurait fixé la température.*

---

* Intuitivement, on devrait s'attendre à ce que H(Pression/Température) $\leq$ H(Pression)
Nous montrerons ceci dans la proposition II.17.

Cette possibilité de "fixer" certaines variables dans un système est mise à profit par l'intermédiaire des *Transinformations A-Conditionnelles*, définies comme suit :

$$\dot{I}_A : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R} \qquad [II.19]$$

$$B \longmapsto \dot{I}_A(B) := \sum_{X \in B} H(X/A) - H(B/A)$$

et

$$\overleftrightarrow{I}_A : \mathcal{P}(\mathcal{P}(\Sigma)) \longrightarrow \mathbb{R} \qquad [II.20]$$

$$R=\{R_1, R_2, \ldots, R_r\} \longmapsto \overleftrightarrow{I}_A(R) := \sum_{i=1}^{r} H(R_i/A) - H(\bigcup_{i=1}^{r} R_i/A)$$

qui mesurent les transinformations lorsque le groupe de variables A est fixé.

REMARQUE : Lorsqu'on conditionne la transinformation entre deux variables avec le reste du système, on obtient la *Transinformation Directe* entre ces deux variables [ASHB 65] :

$$\overleftrightarrow{I}_D(X_i : X_j) := \overleftrightarrow{I}_{\Sigma \setminus \{X_i, X_j\}}(X_i : X_j) \qquad [II.21]$$

$$:= H(X_i/\Sigma\setminus\{X_i,X_j\}) + H(X_j/\Sigma\setminus\{X_i,X_j\}) - H(X_i,X_j/\Sigma\setminus\{X_i,X_j\})$$

$$:= H(\Sigma\setminus\{X_j\}) - H(\Sigma\setminus\{X_i,X_j\}) + H(\Sigma\setminus\{X_i\}) - H(\Sigma\setminus\{X_i,X_j\})$$

$$- H(\Sigma) + H(\Sigma\setminus\{X_i,X_j\})$$

$$= H(\Sigma\setminus\{X_i\}) + H(\Sigma\setminus\{X_j\}) - H(\Sigma\setminus\{X_i,X_j\}) - H(\Sigma).$$

Celle-ci mesure la transmission d'information entre X_i et X_j qui ne dépend que de X_i et X_j, c'est-à-dire, qui ne passe pas par l'intermédiaire des autres variables.

Nous reviendrons au paragraphe § 8.1 sur les transinformations conditionnelles. Pour l'instant, nous allons étendre ce concept de conditionnement à des semivaluations autres que l'entropie.

6.2 - CAS H=SEMI-VALUATION QUELCONQUE : SEMI-VALUATION ET INTERVALUATION CONDITIONNELLES

Dans le paragraphe précédent, [II.11] définit l'entropie A-Conditionnelle. Lorsque H est une semi-valuation quelconque on définit de façon identique une *Semi-Valuation Conditionnelle* H(./A). Les relations [II.12] à [II.16] restent vérifiées dans ce cas; de même, [II.19] et [II.20] définissent les *Intervaluations Conditionnelles* $\dot{I}_A$ et $\overleftrightarrow{I}_A$. Voyons maintenant quelques propriétés :

Proposition II.17 (SV) Quel que soit $A \in \mathcal{P}(\Sigma)$:

$$\forall\, B \in \mathcal{P}(\Sigma) : H(B/A) \leqslant H(B)$$

Interprétation : on a noté dans l'exemple du § 6.1 qu'il était raisonnable d'espérer une diminution de l'entropie de la variable "Pression" lorsque la "température" était fixée. La proposition ci-dessus dit que la connaissance d'un groupe quelconque de variables ne peut que réduire l'incertitude attachée à n'importe quel autre groupe. Par ailleurs, cette réduction n'est nulle que si les deux groupes de variables sont indépendants.

Démonstration :
==========

$$H(A \cup B) \leqslant H(A) + H(B)$$

$$\Rightarrow \quad H(B/A) := H(A \cup B) - H(A) \leqslant H(B)$$

(F.D.)

Proposition II.18 (SVM) Quel que soit $A \in \mathcal{P}(\Sigma)$:
---

① $\forall\, B \in \mathcal{P}(\Sigma)$: $H(B/A) \geqslant 0$

② $\forall\, B, C \in \mathcal{P}(\Sigma)$: $B \subseteq C \Rightarrow H(B/A) \leqslant H(C/A)$.

Interprétation :
==========

L'affirmation ① dit que la quantité d'information qui nous reste à connaître est toujours non négative (autrement dit, il n'y a jamais un "excédent" d'information connue). L'affirmation ② dit que si H est monotone, alors H(./A) l'est aussi.

Démonstration :
==========

① Quels que soient $A,B \in \mathcal{P}(\Sigma)$

$A \subseteq A \cup B$

$\Rightarrow H(A) \leqslant H(A \cup B)$

$\Rightarrow 0 \leqslant H(A \cup B) - H(A) =: H(B/A)$.

② $B \subseteq C \Rightarrow A \cup B \subseteq A \cup C$

$\Rightarrow H(A \cup B) \leqslant H(A \cup C)$

$\Rightarrow H(B/A) := H(A \cup B) - H(A) \leqslant H(A \cup C) - H(A) =: H(C/A)$.

(F.D.)

REMARQUE : Cette dernière proposition fait intervenir l'hypothèse de monotonicité de la semi-valuation H dont il est question. En particulier la démonstration de l'affirmation (1) dit que cette hypothèse est requise pour que la semi-valuation A-conditionnelle soit positive. Cela nous laisse penser qu'il y a des semi-valuations pour lesquelles la semi-valuation conditionnelle obtenue n'est pas toujours positive. Mais étant donné qu'on avait montré que toute semi-valuation est positive (c.f. proposition II.6 (SV)) on se pose la question : Est-ce qu'une semi-valuation conditionnelle est une semi-valuation ? Autrement dit, si A, B, C sont des éléments quelconques de $\mathcal{P}(\Sigma)$, est-il toujours vrai, quelle que soit la semi-valuation H, que

$$H(B \cup C/A) \leqslant H(B/A) + H(C/A) \ ?$$

Nous allons répondre à cette question au paragraphe suivant.

7 - LE CONCEPT DE SEMI-MODULATION

7.1 - SEMI-MODULATION ET ENTROPIE

L'application

$$H : \mathcal{P}(\Sigma) \longrightarrow \mathbb{R}$$
$$A \longmapsto H(A)$$

est appellée une *semi-Modulation* (SM) ssi elle vérifie

$$\forall A,B \in \mathcal{P}(\Sigma) : H(A \cup B) + H(A \cap B) \leqslant H(A) + H(B). \qquad [II.22]$$

H est appellée une *Semi-Modulation Positive* (SMP) ssi en plus de [II.22] elle vérifie

$$\forall A \in \mathcal{P}(\Sigma) : \quad H(A) \geqslant 0 \qquad [II.23]$$

H est appellée une *Semi-Modulation Positive Monotone* (SMPM) ssi en plus de [II.22] et [II.23], elle vérifie

$$\forall A,B \in \mathcal{P}(\Sigma) : A \subseteq B \Rightarrow H(A) \leqslant H(B) \qquad [II.24]$$

On peut remarquer dès maintenant que toute semi-modulation positive (SMP) est une semi-valuation (SV). En effet
$H(A \cup B) \leqslant H(A \cup B) + H(A \cap B) \leqslant H(A) + H(B)$.

En particulier donc, toute SMPM est une SVM.

Proposition II.19

L'entropie est une Semi-Modulation Positive Monotone.

Positive Monotone

Démonstration :

On sait déjà que l'entropie est positive et monotone. Il nous reste à montrer que

$$\forall A,B \in \mathcal{P}(\Sigma) : H(A\cup B) + H(A\cap B) \leq H(A) + H(B) \qquad [II.22]$$

Si $A\cap B = \emptyset$, cette inégalité devient la condition de semi-valuation, dont on sait déjà qu'elle est vérifiée par l'entropie.

Considérons donc le cas $A\cap B \neq \emptyset$. Soient alors

$$X = A\cap B,\ Y = A\setminus B\ ,\ Z = B\setminus A$$

les variables vectorielles dont les composantes sont les éléments de l'ensemble indiqué. Avec ces notations, l'inégalité [II.22] s'écrit :

Quels que soient X,Y,Z : $H(X,Y,Z) + H(X) \leq H(X,Y) + H(X,Z)$ [II.25]

Nous allons montrer que $H(X,Y) + H(X,Z) - H(X) - H(X,Y,Z) \geq 0$. Soient

$$P_{ijK} := Pr(X = x_i,\ Y = y_j,\ Z = z_K)$$

$$P_{.jK} := Pr(Y = y_j,\ Z = z_K) = \sum_i P_{ijK}$$

$$P_{i.K} := Pr(X = X_i,\ Z = Z_K) = \sum_j P_{ijK}$$

$$P_{..K} := Pr(Z = z_K) = \sum_{ij} P_{ijk}\ ,\ \text{etc}\ .$$

$$H(X,Y) + H(X,Z) - H(X) - H(X,Y,Z)$$

$$:= -\sum_{ij} P_{ij.}\log P_{ij.} - \sum_{iK} P_{i.K}\log P_{i.K} + \sum_i P_{i..}\log P_{i..} + \sum_{ijK} P_{ijK}\log P_{ijK}$$

$$=-\sum_{ij}\Big(\sum_{K} P_{ijK}\Big)\cdot\log P_{ij.} -\sum_{iK}\Big(\sum_{j} P_{ijK}\Big)\cdot\log P_{i.K}+\sum_{i}\Big(\sum_{jK} P_{ijK}\Big)\cdot\log P_{i..} + \sum_{ijK} P_{ijK}\cdot\log P_{ijK}$$

$$=-\sum_{ijK} P_{ijK}\cdot\log P_{ij.} - \sum_{ijK} P_{ijK}\cdot\log P_{i.K} + \sum_{ijK} P_{ijK}\cdot\log P_{i..} + \sum_{ijK} P_{ijK}\cdot\log P_{ijK}$$

$$=-\sum_{ijK} P_{ijK}\cdot\log \frac{P_{ij.}}{P_{i..}} + \sum_{ijK} P_{ijK}\cdot\log \frac{P_{ijK}}{P_{i.K}}$$

$$=-\sum_{ijK} P_{i.K}\Big(\frac{P_{ijK}}{P_{i.K}}\Big)\cdot\log \frac{P_{ij.}}{P_{i..}} + \sum_{ijK} P_{ijK}\cdot\log \frac{P_{ijK}}{P_{i.K}}$$

$$= \sum_{iK} P_{i.K}\left[-\sum_{j}\frac{P_{ijK}}{P_{i.K}}\cdot\log \frac{P_{ij.}}{P_{i..}}\right] + \sum_{ijK} P_{ijK}\cdot\log \frac{P_{ijK}}{P_{i.K}} \qquad [II.26]$$

Voyons maintenant que, pour tout couple i,K :

$$* \frac{P_{ijK}}{P_{i.K}} \geq 0 \,,\, \frac{P_{ij.}}{P_{i..}} \geq 0 \qquad \forall\, j$$

$$* \frac{P_{ij.}}{P_{i..}} = 0 \Rightarrow \frac{P_{ijK}}{P_{i.K}} = 0 \qquad \forall\, j$$

En effet $\frac{P_{ij.}}{P_{i..}} = 0 \Rightarrow P_{ij.} = 0 \Rightarrow P_{ijK} = 0 \Rightarrow \frac{P_{ijK}}{P_{i.K}} = 0$.

$$* \sum_{j} \frac{P_{ijK}}{P_{i.K}} = 1 \quad , \quad \sum_{j} \frac{P_{ij.}}{P_{i..}} = 1$$

ainsi, les hypothèses du Lemme II.1 sont vérifiées; alors pour chaque i,K :

$$-\sum_{j}\frac{P_{ijK}}{P_{i.K}}\cdot\log \frac{P_{ij.}}{P_{i..}} \geq -\sum_{j}\frac{P_{ijK}}{P_{i.K}}\cdot\log \frac{P_{ijK}}{P_{i.K}}$$

En appliquant cette inégalité dans [II.26] on obtient :

$$[II.26] \geq \sum_{iK} P_{i.K} \left[-\sum_{j} \frac{P_{ijK}}{P_{i.K}} \cdot \log \frac{P_{ijK}}{P_{i.K}} \right] + \sum_{ijK} P_{ijK} \cdot \log \frac{P_{ijK}}{P_{i.K}} = 0$$

(F.D.)

7.2. - RELATION ENTRE LES CONCEPTS DE SEMI-VALUATION ET SEMI-MODULATION

Avant de présenter quelques propriétés sur les semi-modulations positives, faisons brièvement le point sur la relation entre les concepts SV, SVM, SMP, SMPM.

Comme nous l'avons montré,

- si H est une SMP $\Rightarrow$ H est une SV

et trivialement

- si H est une SMPM $\Rightarrow$ H est une SVM,
- si H est une SMPM $\Rightarrow$ H est une SMP,
- si H est une SVM $\Rightarrow$ H est une SV.

On peut visualiser ces implications dans la figure II.3. présentée dans la page suivante.

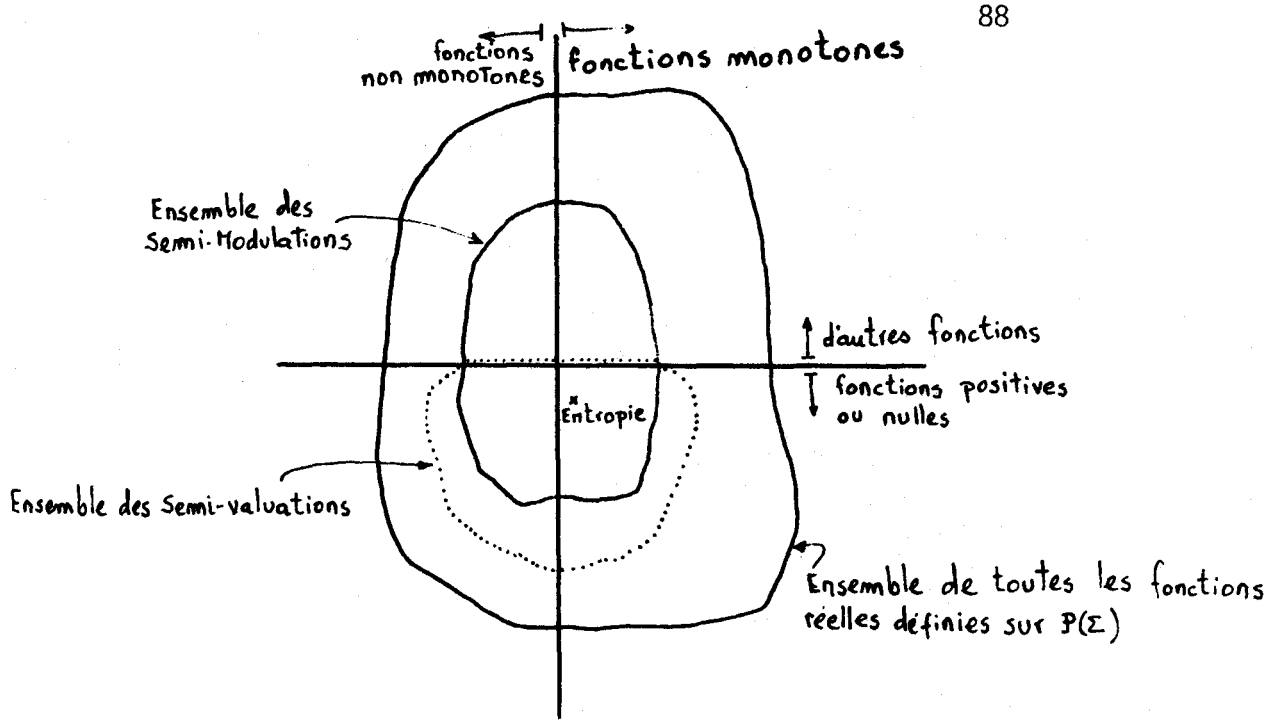


Figure II.3 - L'ensemble des fonctions réelles définies sur $\mathcal{P}(\Sigma)$

Ainsi donc, on peut dire que :

- un résultat valable pour toute SV, est en particulier valable pour toute SVM, SMP, SMPM.
- un résultat valable pour toute SVM, est en particulier valable pour toute SMPM.
- un résultat valable pour toute SMP, est en particulier valable pour toute SMPM.

7.3 - PROPRIETES DES SEMI-MODULATIONS

La propriété de semi-modulation positive nous permet d'établir quelques propositions peu connues, et en tout cas, qui n'avaient été montrées que pour H=entropie.

Proposition II.20 (SMP)

Si H est une SMP, alors

$$\forall A,B \in \mathcal{P}(\Sigma) : H(A \cap B) \leq \overleftrightarrow{I}(A:B)$$

Interprétation : pour H=entropie, l'explication est facile : l'information qu'on trouve en même temps en A et en B doit au moins être égale à celle des variables qui font partie de A et de B (variables de couplage).

Démonstration :

H est une SMP

$$\forall A, B \in \mathcal{P}(\Sigma): H(A \cup B) + H(A \cap B) \leq H(A) + H(B)$$

$$\Rightarrow H(A \cap B) \leq H(A) + H(B) - H(A \cup B) =: \overleftrightarrow{I}(A:B)$$

(F.D.)

Proposition II.21 (SMP)

Si H est une SMP, alors

$$\forall A,B,C \in \mathcal{P}(\Sigma): H(B \cup C/A) + H(B \cap C/A) \leq H(B/A) + H(C/A)$$

Cette affirmation dit que si H est une SMP, alors $H(\cdot/A)$ est aussi une SMP. Cela implique que $H(\cdot/A)$ est une SV.

Démonstration :

La condition de semi-modulation implique que

$$\forall A,B,C \in \mathcal{P}(\Sigma) : H((AuB)u(AuC)) + H((AuB)n(AuC)) \leqslant H(AuB) + H(AuC)$$

$$\Rightarrow \quad H(AuBuC) + H((BnC)uA) \leqslant H(AuB) + H(AuC)$$

$$\Rightarrow \quad H(AuBuC) - H(A) + H((BnC)uA) - H(A) \leqslant H(AuB) - H(A) + H(AuC) - H(A)$$

$$\Rightarrow \quad H(BuC/A) + H(BnC/A) \leqslant H(B/A) + H(C/A)$$

(F.D.)

Proposition II.22 (SMPM)

Si H est une SMPM alors

$$\forall A',A \in \mathcal{P}(\Sigma) : A' \subseteq A \Rightarrow \left[\forall B \in \mathcal{P}(\Sigma) : H(B/A') \geqslant H(B/A) \right]$$

Interprétation : pour H=entropie, elle est évidente : plus l'ensemble des variables déjà connues est grand, moindre sera l'incertitude sur n'importe quelle autre partie B.

Démonstration :

$$A' \subseteq A'u(BnA) = (A'uB) \; n \; (A'uA)$$

$$\Rightarrow \quad H(A') \leqslant H((A'uB)n(A'uA))$$

alors

$H(A') + H((A' \cup B) \cup (A' \cup A))$

$\leq H((A' \cup B) \cap (A' \cup A)) + H((A' \cup B) \cup (A \cup A'))$, et d'après la propriété de SM

$\leq H(A' \cup B) + H(A' \cup A)$

ainsi

$H(A') + H((A' \cup B) \cup (A' \cup A)) \leq H(A' \cup B) + H(A' \cup A)$

$\Rightarrow H((A' \cup B) \cup (A' \cup A)) - H(A' \cup A) \leq H(A' \cup B) - H(A')$

Etant donné que $(A' \cup B) \cup (A' \cup A) = A' \cup B \cup A = B \cup A$, et $A' \cup A = A$,

$\Rightarrow H(B/A) := H(B \cup A) - H(A) \leq H(A' \cup B) - H(A') =: H(B/A')$

(F.D.)

Proposition II.23 (SMPM)

Si H est une SMPM alors

$\forall A, A', B, B' \in \mathcal{P}(\Sigma) : A' \subseteq A,\ B' \subseteq B \quad \overleftrightarrow{I}(A' : B') \leq \overleftrightarrow{I}(A : B)$

Interprétation : pour H=entropie, cette proposition affirme que si les groupes de variables deviennent plus grands, leur échange d'information ne peut qu'augmenter.

Démonstration :
==========

Si $B' \subseteq B$ alors $H(A/B) \leq H(A/B')$ (c.f. proposition II.22)

D'autre part :

$\overleftrightarrow{I}(A:B) = H(A) - H(A/B)$ d'après [II.15]

$\geqslant H(A) - H(A/B') = \overleftrightarrow{I}(A:B')$

Pareil, $\overleftrightarrow{I}(A:B') = H(B') - H(B'/A)$

$\geqslant H(B') - H(B'/A') = \overleftrightarrow{I}(A':B')$

et ainsi $\overleftrightarrow{I}(A:B) \geqslant \overleftrightarrow{I}(A:B') \geqslant \overleftrightarrow{I}(A':B')$

(F.D.)

Nous pouvons généraliser la proposition ci-dessus ainsi :

Proposition II.24 (SMPM)

Si H est une SMPM et si

$R = \{R_1, R_2, \ldots R_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$.

$R' = \{R'_1, R'_2, \ldots, R'_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$.

avec $\forall_i : R'_i \subseteq R_i$

alors $\overleftrightarrow{I}(R') = \overleftrightarrow{I}(R'_1:R'_2:\ldots:R'_r) \leqslant \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = \overleftrightarrow{I}(R)$

Démonstration :
==========

D'après la proposition II.12, (2) :

$$\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_r)$$

$$= \overleftrightarrow{I}(R_1:R_2) + \overleftrightarrow{I}(R_1uR_2:R_3) + \overleftrightarrow{I}(R_1uR_2uR_3:R_4) + \ldots + \overleftrightarrow{I}(R_1u\ldots uR_{r-1}:R_r)$$

et par application de la proposition II.23 :

$$\geqslant \overleftrightarrow{I}(R'_1:R'_2) + \overleftrightarrow{I}(R'_1uR'_2:R'_3) + \overleftrightarrow{I}(R'_1uR'_2uR'_3:R'_4) + \ldots + \overleftrightarrow{I}(R'_1u\ldots uR'_{r-1}:R'_r)$$

$$= \overleftrightarrow{I}(R'_1:R'_2 : \ldots : R'_r)$$

$$= \overleftrightarrow{I}(R')$$

(F.D.)

La proposition II.24 implique immédiatement deux affirmations qui seront par la suite importantes [STAR 81]

Proposition II.25 (SMPM)

Si $R = \{R_1,R_2,\ldots,R_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$, on a :

① Si $\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) \leqslant \varepsilon$

alors

quel que soit $R'=\{R'_1,R'_2,\ldots,R'_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$, avec $\forall_i : R'_i \subseteq R_i$:

$$\overleftrightarrow{I}(R') = \overleftrightarrow{I}(R'_1:R'_2:\ldots:R'_r) \leqslant \varepsilon$$

② Si $\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = 0$,

alors

quel que soit $R'=\{R'_1,R'_2,\ldots,R'_r\}$ avec $\forall i : R'_i \subseteq R_i$:

$$\overleftrightarrow{I}(R') = \overleftrightarrow{I}(R'_1:R'_2: \ldots : R'_r) = 0 .$$

Proposition II.26 (SMPM)

Supposons que $\forall X_i \in \Sigma : H(X_i) > 0$.
Alors quel que soit $R=\{R_1,R_2,\ldots,R_r\} \in \mathcal{P}(\mathcal{P}(\Sigma))$:

$$\overleftrightarrow{I}(R) = \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = 0 \Rightarrow \forall i \neq j : R_i \cap R_j = \emptyset$$

Interprétation : pour H=entropie, cette proposition dit que si le système Σ ne comporte pas de variables constantes, alors un échange nul d'information entre quelques groupes de variables ne peut arriver que si les groupes de variables sont disjoints .

Démonstration :

D'après la proposition II.9, (2) :

$$\overleftrightarrow{I}(R_1:R_2:\ldots:R_r) = 0 \Rightarrow \forall i \neq j : \overleftrightarrow{I}(R_i:R_j) = 0$$

Mais d'après la proposition II.20 :

$0 \leq H(R_i \cap R_j) \leq \overleftrightarrow{I}(R_i:R_j) = 0$, donc $H(R_i \cap R_j) = 0$.

D'autre part, vu que H est monotone, le fait que

$\forall X_i \in \Sigma : H(X_i) > 0$ implique que $\forall S \in \mathcal{P}(\Sigma), S \neq \emptyset : H(S) > 0$.

Ainsi donc, $H(R_i \cap R_j) = 0$ implique forcément que $R_i \cap R_j = \emptyset$.

(F.D.)

8 - LE PRINCIPE DE CONDITIONNEMENT UNIFORME

8.1 - VERSION ORIGINALE POUR H=ENTROPIE

Au paragraphe § 6.1 nous avons introduit les concepts d'Entropie et Transinformation Conditionnelles. Pour manipuler ces concepts on dispose du *Principe de Conditionnement Uniforme* (appelé aussi Sous-Indexation Uniforme). Ce principe, que W.R. ASHBY [1965] attribue à W.R. GARNER [1956] et W.H. MCGILL [1954], dit ceci :

" Si l'on conditionne avec le même ensemble de variables, tous les termes d'une identité valable pour H=Entropie, qui contient $H(\ ), H(\ /.), \overleftrightarrow{I}, \overleftrightarrow{I}_\alpha, \dot{I}, \dot{I}_\beta$ on obtient à nouveau une identité valable pour l'entropie (le double conditionnement étant équivalent au conditionnement avec l'union)".

Ainsi par exemple, à partir des affirmations déjà montrées pour l'entropie :

i) $\forall\ A,B \in \mathcal{P}(\Sigma) : H(AuB) \leqslant H(A) + H(B)$

ii) $\forall\ A,B \in \mathcal{P}(\Sigma) : \overleftrightarrow{I}(A:B) = H(A) - H(A/B)$

iii) $\forall R = \{R_1,R_2,\ldots,R_r\} \in \mathbb{P}(\Sigma) : \overleftrightarrow{I}(R_1:R_2:\ldots:R_r) + \sum_{i=1}^{r} \dot{I}(R_i) = \overleftrightarrow{I}(X_1:X_2:\ldots:X_N)$

nous pourrions obtenir directement, quel que soit $c \in \mathcal{P}(\Sigma)$:

i) $\forall A,B \in \mathcal{P}(\Sigma) : H(AuB/C) \leqslant H(A/C) + H(B/C)$

ii) $\forall A,B \in \mathcal{P}(\Sigma) : \overleftrightarrow{I}_C(A:B) = H(A/C) - H(A/BuC)$

iii) $\forall R = \{R_1,R_2,\ldots,R_r\} \in \mathbb{P}(\Sigma)$:

$$\overleftrightarrow{I}_C(R_1:R_2:\ldots R_r) + \sum_{i=1}^{r} \dot{I}_C(R_i) = \overleftrightarrow{I}_C(X_1:X_2:\ldots:X_N).$$

8.2 - VERSION GENERALISEE POUR H=SMP QUELCONQUE

Avec les propositions que nous avons présentées aux paragraphes précédents, nous sommes en mesure d'étendre le principe de conditionnement uniforme au cadre des semi-valuations et semi-modulations.

En effet, nous avons montré dans la proposition II.21 que si H(.) était une SMP (respec.SMPM), alors H(./A) était aussi une SMP (respect.SMPM), donc une SV (respec.SVM). Ainsi donc, pour H étant une SMP (Respect.SMPM), si dans une identité valable pour toute SMP (respec.SMPM) au lieu de H(.) on met H(./A), l'identité doit rester tout à fait valable.

On peut se poser la question suivante : et si l'identité en question contient déjà des termes conditionnés, comme H(B/C), $\overleftrightarrow{I}_c$, $\dot{I}_c$? Voyons que dans ce cas il suffit de conditionner avec l'union. Par définition:

i) $H(B/C) := H(B \cup C) - H(C)$

ii) $\overleftrightarrow{I}_c(R_1 : R_2 : \ldots : R_r) := \sum_i H(R_i/C) - H(\bigcup_i R_i/C)$

iii) $\dot{I}_c(B) := \sum_{X \in B} H(X/C) - H(B/C)$

Si on reconditionne les deux côtés de ces égalités avec l'ensemble A, on obtient :

i') $H((B/C)/A) = H(B \cup C/A) - H(C/A)$

$:= H(B \cup C \cup A) - H(A) - H(C \cup A) + H(A)$

$= H(B \cup C \cup A) - H(C \cup A)$

$=: H(B/C \cup A)$

$$ii') \quad \overleftrightarrow{I}_{C_A}(R_1:R_2:\ldots:R_r) = \sum_i H((R_i/C)/A) - H((\underset{i}{UR_i}/C)/A)$$

$$= \sum_i H(R_i/C_uA) - H(\underset{i}{UR_i}/CuA)$$

$$=: \overleftrightarrow{I}_{CuA}(R_1:R_2:\ldots:R_r)$$

$$iii') \quad \dot{I}_{C_A}(B) = \sum_{X\varepsilon B} H((X/C)/A) - H((B/C)/A)$$

$$= \sum_{X\varepsilon B} H(X/CuA) - H(B/CuA)$$

$$=: \dot{I}_{CuA}(B).$$

Le *Principe de Conditionnement Uniforme Généralisé* peut alors être énoncé ainsi :

" Quelle que soit H une SMP (respec.SMPM), si dans une identité valable pour toute SMP (respec.SMPM) on conditionne chaque terme avec un même ensemble $A\varepsilon\mathcal{P}(\Sigma)$, on obtient une identité. Le conditionnement se fait ainsi :

- au lieu de $H(.)$ mettre $H(./A)$,
- au lieu de $H(./\alpha)$ mettre $H(./\alpha uA)$,
- au lieu de $\overleftrightarrow{I}(R_1:..:R_r)$ mettre $\overleftrightarrow{I}_A(R_1:....:R_r)$,
- au lieu de $\overleftrightarrow{I}_\alpha(R_1:....:R_r)$ mettre $\overleftrightarrow{I}_{\alpha uA}(R_1:....:R_r)$,
- au lieu de $\dot{I}(.)$ mettre $\dot{I}_A(.)$,
- au lieu de $\dot{I}_\alpha(.)$ mettre $\dot{I}_{\alpha uA}(.)$ "

9 - INTERPRÉTATION GRAPHIQUE

9.1 - L'ANALOGIE ENTRE SEMI-VALUATION ET MESURE

Il est possible d'établir une analogie entre l'entropie et la théorie de la Mesure, qui permet de faire, avec quelques réserves, une interprétation graphique [MILL 63]. Nous allons justifier cette analogie, l'étendre à toute SVM, et montrer ses limitations .

Soit $\mu:\mathfrak{M} \longrightarrow \mathbb{R}$ une mesure [BART 66, chap.I et II] définie sur un ensemble $\mathfrak{M}$ d'objets mesurables (par exemple, la mesure "surface" définie sur l'ensemble des figures planes; ou bien la mesure "volume" définie sur l'ensemble des solides, etc .). On a les deux listes de propriétés :

$H:\mathcal{P}(\Sigma) \longrightarrow \mathbb{R}$	$\mu:\mathfrak{M} \longrightarrow \mathbb{R}$
1) $\forall S \in \mathcal{P}(\Sigma) : H(S) \geq 0$	1) $\forall F \in \mathfrak{M} : \mu(F) \geq 0$
2) $\forall S_1, S_2 \in \mathcal{P}(\Sigma) : H(S_1 \cup S_2) \leq H(S_1)+H(S_2)$	2) $\forall F,G \in \mathfrak{M} : \mu(F//G) \leq \mu(F)+\mu(G)$, où // veut dire "concatené","réuni", "assemblé".
3) $H(S_1 \cup S_2)=H(S_1) + H(S_2) :\Longleftrightarrow$ S_1 et S_2 sont H-indépendants (s'il n'y a aucune information qui se trouve aussi bien en A qu'en B), ou si $H(S_1) = 0$ ou $H(S_2) = 0$	3) $\mu(F//G) = \mu(F) + \mu(G) \Longleftrightarrow$ F et G sont disjoints, ou si $\mu(F) = 0$ ou $\mu(G) = 0$
4) $S_1 \subseteq S_2 \Rightarrow H(S_1) \leq H(S_2)$	4) F "est un morceau de" G $\Rightarrow$ $\mu(F) \leq \mu(G)$

Ce parallèle évident nous permet de faire l'interprétation graphique suivante :

- chaque variable X_i sera représentée par un cercle (ou figure quelconque) dont la surface est proportionnelle à $H(X_i)$.

- la surface d'intersection entre les cercles X_i et X_j sera proportionnelle à $\overleftrightarrow{I}(X_i:X_j)$.

Par exemple pour deux variables X et Y :

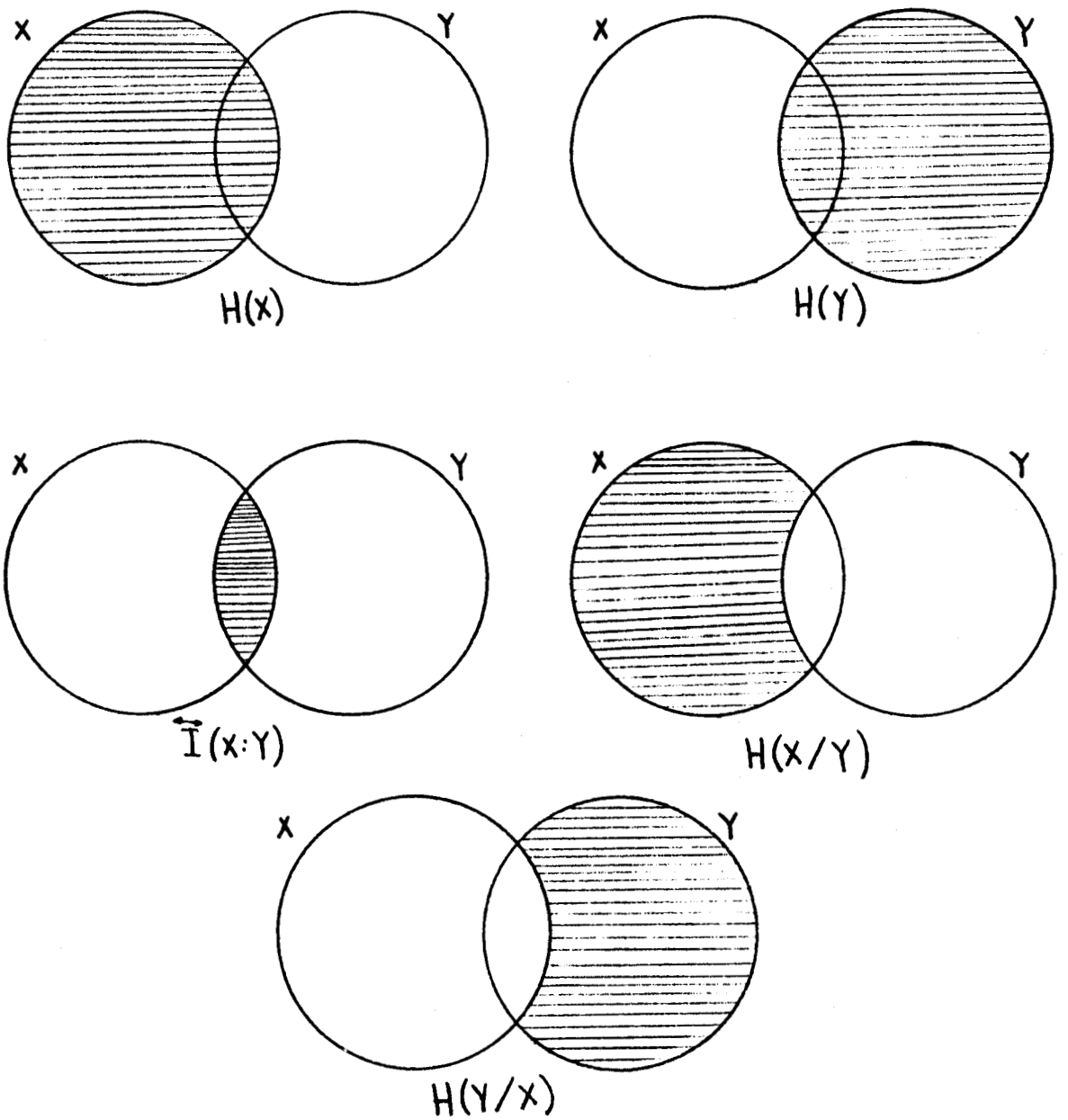


Figure II.4 : Représentation graphique de H : deux variables

Pour trois variables on a :

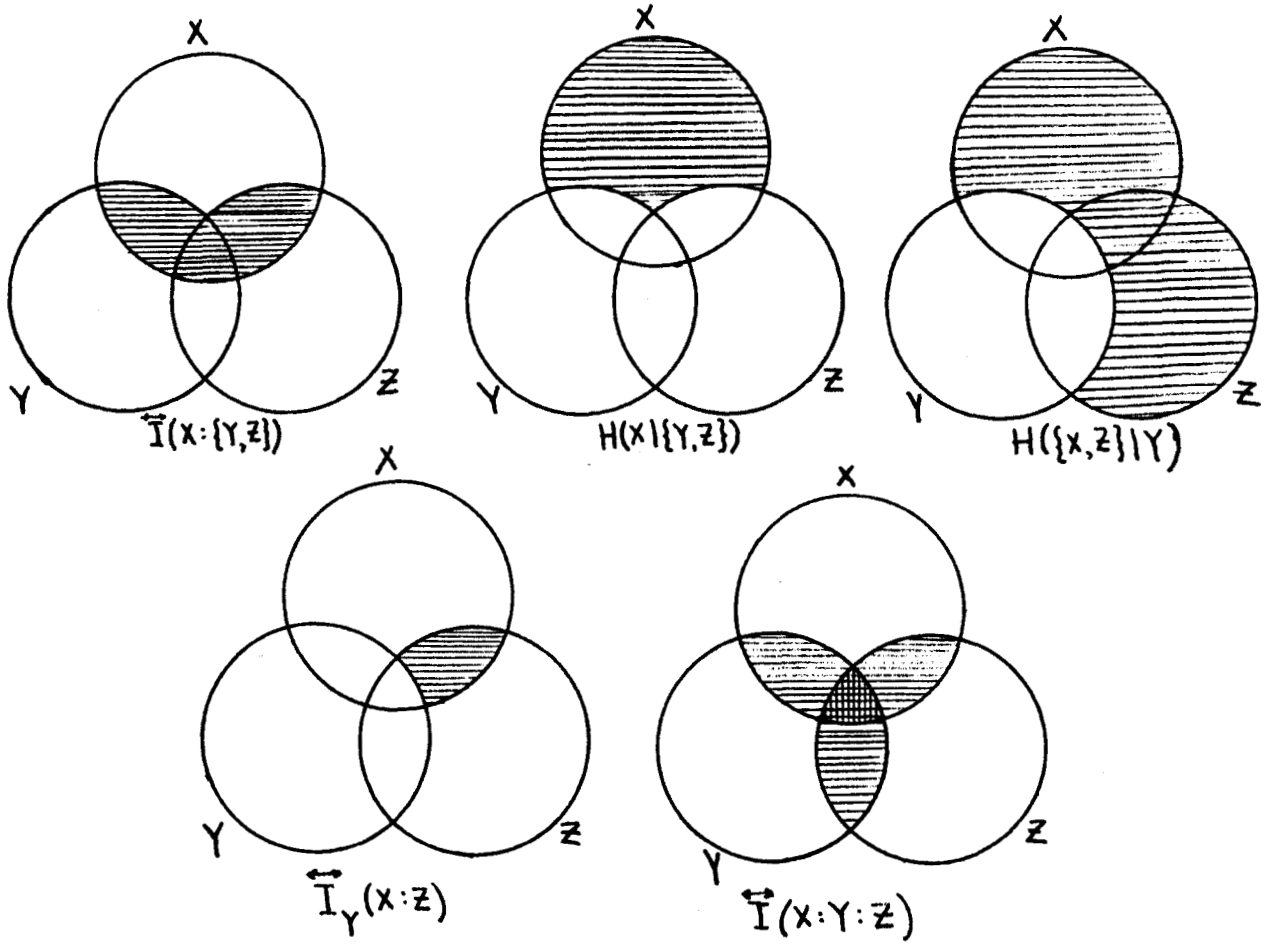


Figure II.5 : Représentation graphique de H : trois variables

Dans ce dernier dessin, le calcul montre que le morceau ▦, commun à 3 variables, doit être compté deux fois. En général, lorsqu'on calcule l'intervaluation entre n variables, la partie du dessin correspondant à l'intersection de K d'entre elles ($2 \leqslant K \leqslant n$), compte K-1 fois.

Vu la similitude entre les propriétés d'une SVM et d'une mesure de surface, on serait tenté de dire que l'analogie est totale. Mais en réalité il y a une subtile différence :

En théorie de la Mesure on peut dire :

F,G disjoints, donc $\mu(F//G) = \mu(F) + \mu(G)$, et
G,K disjoints, donc $\mu(G//K) = \mu(G) + \mu(K)$, et
F,K disjoints, donc $\mu\ (F//K) = \mu(F) + \mu(K)$,

implique

F,G,K sont disjoints, donc $\mu(F//G//K) = \mu(F) + \mu(G) + \mu(K)$.

Mais avec les SVM cela n'est pas forcément vrai :

X,Y H-indépendants, donc H(X,Y) = H(X) + H(Y), et
Y,Z H-indépendants, donc H(Y,Z) = H(Y) + H(Z), et
X,Z H-indépendants, donc H(X,Z) = H(X) + H(Z),

n'implique pas (c.f. §.9.3)

X,Y,Z H-indépendants, donc H(X,Y,Z) = H(X) + H(Y) + H(Z)

Etant donné que l'analogie n'est pas totale, on peut se poser les questions suivantes : Est-ce que tout système réel possède une représentation graphique ?; Est-ce que toute représentation graphique correspond à un système réel ? les paragraphes suivants répondent à cette question pour H=Entropie.

9.2 - A TOUTE REPRESENTATION GRAPHIQUE CORRESPOND UN SYSTEME "REEL"

Pour montrer ceci, pour une représentation graphique donnée nous allons faire la synthèse d'un système ne contenant que des variables dichotomiques. Voyons tout d'abord la
a- construction d'une variable ayant une entropie donnée :

La fonction h(p) := p.log p - (1-p).log (1-p), avec $0 \leqslant p \leqslant 1$, indique l'entropie d'une variable binaire b qui prendrait les modalités 0 et 1 avec les probabilités p et 1-p (fig.II.6).

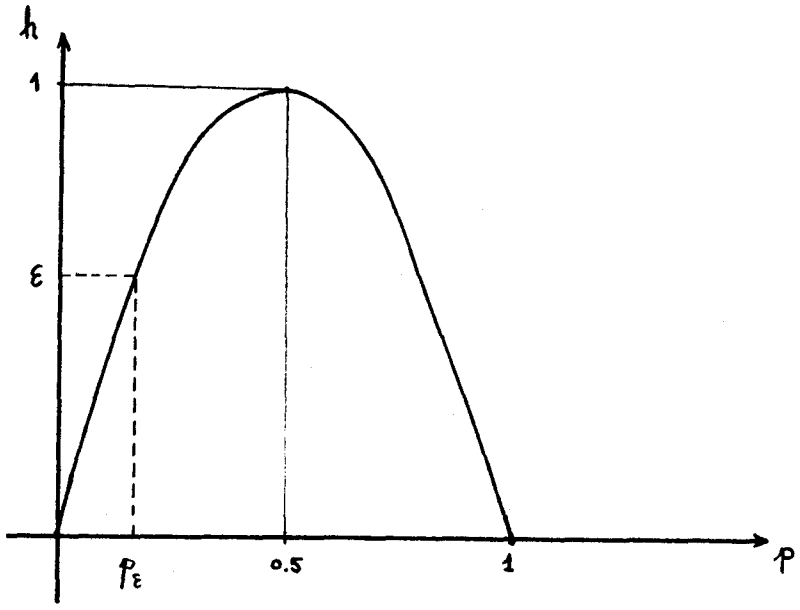


Figure II.6 - Entropie d'une variable dichotomique b

On remarque toute de suite que si la variable binaire b prend les valeurs 0 et 1 avec la même probabilité 1/2, alors $H(b) = h(1/2) = 1$. Il est clair aussi que quelque soit ε, $0 \leqslant \varepsilon \leqslant 1$, il existe p_ε, $0 \leqslant p_\varepsilon \leqslant 1/2$, tel que $h(p_\varepsilon) = \varepsilon$.

Voyons donc que $\forall\ \alpha \in \mathbb{R}$, $\alpha > 0$, il est possible de construire une variable X avec $H(X) = \alpha$. Pour cela on peut procéder ainsi :

- Si α est un entier : on définit

$$X := b_1 b_2 b_3 \ldots b_\alpha$$

où les b_i sont des variables binaires *indépendantes*, qui prennent la valeur 0 ou 1 avec la même probabilité 1/2. Alors

$$H(X) = H(b_1 b_2 \ldots b_\alpha) = \underbrace{1 + 1 + \ldots + 1}_{\alpha} = \alpha$$

- Si α n'est pas un entier : Soit $[\alpha]$ la partie entière de α, et $\langle\alpha\rangle := \alpha - [\alpha]$ la partie décimale (bien entendu, $0 < \langle\alpha\rangle < 1$). On définit :

$$X := b_1 b_2 \cdots b_{[\alpha]} \tilde{b}$$

où, comme auparavant, $b_1 b_2 \ldots b_{[\alpha]}$ sont $[\alpha]$ variables binaires indépendantes qui prennent la valeur 0 ou 1 avec la même probabilité 1/2. On sait par ailleurs qu'il existe $p_{\langle\alpha\rangle}$ tel que $h(p_{\langle\alpha\rangle}) = \langle\alpha\rangle$. Alors on définit $\tilde{b}$ par une variable binaire, indépendante de $b_1 \ldots b_{[\alpha]}$, qui prend les valeurs 0 et 1 avec les probabilités $p_{\langle\alpha\rangle}$ et $1-p_{\langle\alpha\rangle}$. Ainsi

$$H(X) = H(b_1 b_2 \ldots b_{[\alpha]} \tilde{b}) = H(b_1) + \ldots + H(b_{[\alpha]}) + H(\tilde{b}) = \underbrace{1+\ldots+1}_{[\alpha]} + \langle\alpha\rangle = \alpha.$$

b- construction des variables du système :

Une représentation graphique de l'entropie est toujours composée d'un certain nombre de *secteurs élémentaires* $\varphi_1, \varphi_2, \varphi_3, \ldots$ correspondants aux intersections des différentes variables (fig.II.7).

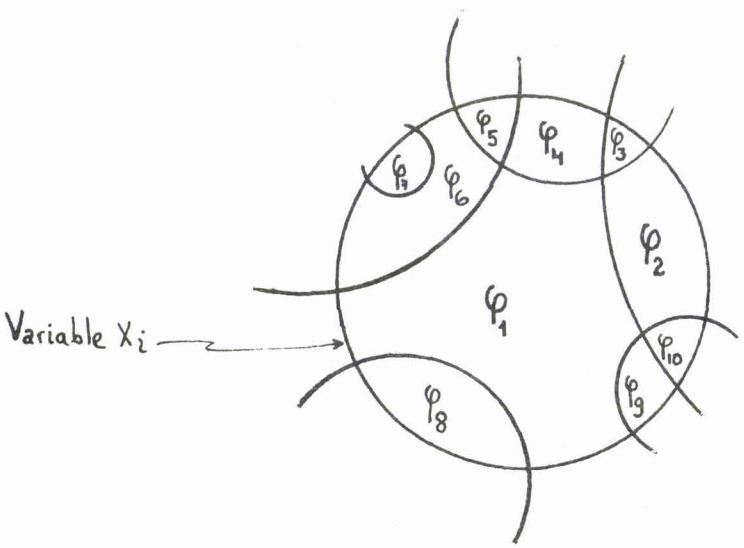


Figure II.7 : Découpage en secteurs élémentaires d'une variable X_i

Pour chaque φ_j on construit une variable X_{φ_j} telle que $H(X_{\varphi_j})$ soit égale à la surface du secteur φ_j. Chaque variable X_i sera alors définie par la concaténation des variables X_{φ_j} telles que φ_j est contenu dans la surface de la variable X_i. Par construction donc chaque variable aura une entropie égale à sa surface et les transinformations entre elles seront égales aux surfaces d'intersection.

9.3 - TOUT SYSTEME REEL N'A PAS DE REPRESENTATION GRAPHIQUE

Nous allons présenter un système de 4 variables binaires pour lequel il n'y a pas de représentation graphique. Ce système est défini par le tableau de données

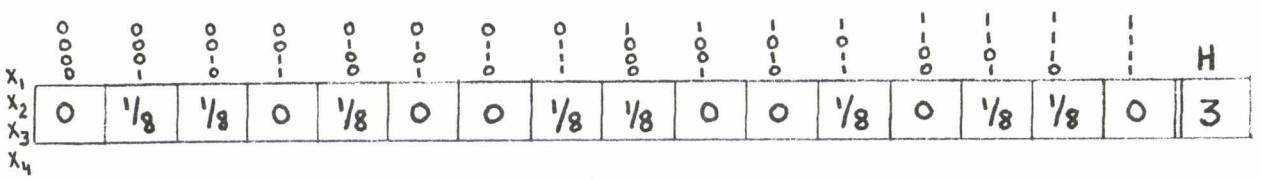
Ω \ Σ	X_1	X_2	X_3	X_4
ω_1	0	1	0	0
ω_2	1	0	1	1
ω_3	1	1	1	0
ω_4	1	1	0	1
ω_5	1	0	0	0
ω_6	0	0	1	0
ω_7	0	1	1	1
ω_8	0	0	0	1

Le calcul des probabilités conjointes et de l'entropie pour tous les sous-ensembles de $\Sigma := \{X_1, X_2, X_3, X_4\}$ donne

	0	1	H
X_1	1/2	1/2	1
X_2	1/2	1/2	1
X_3	1/2	1/2	1
X_4	1/2	1/2	1

	00	01	10	11	H
X_1X_2	1/4	1/4	1/4	1/4	2
X_1X_3	1/4	1/4	1/4	1/4	2
X_1X_4	1/4	1/4	1/4	1/4	2
X_2X_3	1/4	1/4	1/4	1/4	2
X_2X_4	1/4	1/4	1/4	1/4	2
X_3X_4	1/4	1/4	1/4	1/4	2

	000	001	010	011	100	101	110	111	H
$X_1X_2X_3$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	3
$X_1X_2X_4$	18/	1/8	1/8	1/8	1/8	1/8	1/8	1/8	3
$X_1X_3X_4$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	3
$X_2X_3X_4$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8	3



D'après les entropies calculées, on obtient

$$\forall i \neq j : \overleftrightarrow{I}(X_i : X_j) = 0$$

Pour la représentation graphique cela veut dire que les cercles correspondants aux variables ne se coupent pas. Donc, la représentation graphique serait constituée de 4 cercles isolés, chacun de surface 1. Ainsi, une étude ne faisant qu'une analyse des relations

ou couplages entre paires de variables va conclure qu'il n'y a aucun lien entre ces variables. Cela est d'ailleurs confirmé lorsqu'on voit que

$$\forall\ i \neq j \neq K \neq i\ :\ \overleftrightarrow{I}(X_i : X_j : X_k) = 0.$$

Mais par contre, on obtient

$$\overleftrightarrow{I}(X_1 : X_2 : X_3 : X_4) = 1$$

Cela implique qu'il existe une surface égale à 1, commune aux 4 cercles. Ceci est géométriquement impossible, puisqu'on avait trouvé que les 4 cercles étaient disjoints. On doit alors conclure qu'il n'y a pas de représentation graphique possible.

Le fait que $\overleftrightarrow{I}(X_1 : X_2 : X_3 : X_4) = 1$ implique qu'il y a bien un lien entre les variables. Cela devient évident lorsqu'on obtient $H(X_i / X_j, X_K, X_l) = 0$ (i,j,K,l tous différents), ce qui veut dire qu'une variable est complètement déterminée lorsqu'on connait les autres trois variables. En effet, si l'on réorganise le tableau de données, la relation entre les variables apparait très claire :

Ω \ Σ	X_1	X_2	X_3	X_4
ω_8	0	0	0	1
ω_6	0	0	1	0
ω_1	0	1	0	0
ω_7	0	1	1	1
ω_5	1	0	0	0
ω_2	1	0	1	1
ω_4	1	1	0	1
ω_3	1	1	1	0

Figure II.8 Un compteur de 3 bits + bit de parité.

Il s'agit donc d'une séquence binaire de trois bits avec son bit de parité (=0 si nombre impair de 1, = 1 si nombre pair de 1).

Ainsi, ce qui semblait être au début un système de 4 variables indépendantes, n'en contient en réalité que 3, la quatrième étant déterminée univoquement par les 3 autres.

10 - CONCLUSIONS

Nous avons présenté dans ce chapitre un certain nombre de propriétés de l'information circulant dans un système, en essayant de mettre en valeur aussi bien l'interprétation intuitive que les possibilités de généralisation de chaque résultat. Cette double démarche, intuitive et générale, nous suggére quelques remarques.

- l'Entropie et les indices de transinformation qui en découlent formalisent avec une très grande justesse le concept naturel de "Information": " ... no other single measure of the amount of information that a random variable X gives about another random variable Y has all the desirable properties possessed by Shannon's measure $\overleftrightarrow{I}(X:Y)$..." [BLAC 68].

- la plupart des propriétés connues des transinformations découlent du concept très simple de Semi-valuation.

- l'utilisation du concept de Semi-Modulation, plus fort que celui de semi-valuation, nous a permis de caractériser quelques propriétés peu connues de l'entropie et des transinformations, ainsi que de démontrer et généraliser le Principe de Conditionnement Uniforme.

- Enfin, l'exemple du paragraphe précédent met en évidence un principe important : le seul moyen qui permet vraiment d'établir tous les liens et relations existant dans un système de N variables, est de faire une analyse de "dimension N", c'est-à-dire, qui prenne en compte les N variables simultanément. Comme cet exemple le montre, toute autre analyse de dimension inférieure à N peut laisser échapper des renseignements, parfois importants, sur la structure du système. C'est celà qui nous a conduit à développer des méthodes et des algorithmes qui évoluent dans le cadre général de l'ensemble de tous les ensembles (des ensembles) de variables, en faisant une analyse N-dimensionnel. C'est l'atout principal des algorithmes que nous présentons dans le chapitre suivant.

CHAPITRE III

SUR LA DÉCOMPOSITION DES SYSTÈMES COMPLEXES

"Divide·ut·regnes"
Jules CESAR (101-44 av.J.C.)
N.MACHIAVEL (1469-1527)

I - INTRODUCTION

En général, le qualificatif *complexe* est donné aux systèmes constitués d'un nombre élevé d'éléments, qui intéragissent d'une façon non simple [SIMO 62]. Cependant, la complexité n'est pas seulement une caractéristique du système en lui-même, mais elle dépend aussi et pour une partie importante de la relation entre l'observateur et le système [CONA 72]. La motivation de ce chapitre est donc la conviction qu'une Décomposition judicieuse du système peut réduire largement sa complexité apparente.

Le concept de Décomposition de Systèmes Complexes peut être compris de diverses manières, lesquelles ne sont pas exclusives, mais bien au contraire, plutôt complémentaires :

i - Décomposition *Temporelle*, correspondant par exemple aux différentes situations dans lesquelles peut se trouver un système dans le temps; ou bien, selon les différentes dynamiques, rapides ou lentes, de l'évolution des éléments du système [FOSS 81, MAGN 81] ; etc .

ii - Décomposition des *Objectifs*, applicable particulièrement à des systèmes qui présentent une multiplicité d'objectifs, le plus souvent incommensurables et/ou antagonistes [KOUV 76; STAR 79].

iii - Décomposition *Verticale*, où les fonctions de commande sont hiérarchisées selon leur horizon d'action et d'observation [RICH 75, chap.I; DUFO 79, chap.I]. C'est le type de décomposition usuel dans les techniques de contrôle hiérarchisé [SING 78; TITL 79; FIND 80]. On peut distinguer deux formes :

iii.1 - Décomposition par *Niveaux d'Influence :* les activités de commande sont regroupées dans des niveaux qui portent sur un fonctionnement d'autant plus général qu'on s'élève dans la hiérarchie, les décisions des plus hauts niveaux étant prioritaires et moins fréquentes que celles des niveaux inférieurs. C'est par exemple le cas d'une entreprise de production : au niveau inférieur il est question de contrôler la qualité et la quantité de production; à un niveau plus élevé on fixe les plans de production à moyen terme, on fait le choix des matières premières, des fournisseurs, des équipements, ...; au niveau supérieur on établit les grandes lignes à suivre, ainsi que les objectifs sociaux ou économiques à long terme, ...

iii.2 - Décomposition par *Niveaux Fonctionnels :* Similaire à la précédente, mais dans ce cas c'est la définition même du contrôle qui est hiérarchisée [BONA 81; LANG 81; CALV 81] . Toujours avec l'exemple d'une entreprise de production, cette approche donnerait lieu à une hiérarchie où le niveau inférieur (de commande directe) a pour rôle d'assurer le fonctionnement du processus selon certaines valeurs de consigne données, et cela en dépit des perturbations affectant les variables. Le second niveau (d'optimisation) est chargé de la détermination des valeurs de consigne pour les différents processus, basé sur des modèles, critères et contraintes qui lui sont donnés. Le troisième niveau (de modélisation) établit les modèles, critères, intéractions des différents processus, en tenant compte de la structure, des ressources et des objectifs fixés. Enfin, le niveau supérieur détermine la structure, les ressources et les objectifs du système global.

iv - Décomposition sur la base de la *Partition de l'ensemble des variables du système,* qui établit un regroupement conceptuel des variables en sous-systèmes faiblement couplés. C'est cette forme de décomposition qui sera l'objet d'étude de ce chapitre.

De nombreuses argumentations ont été données pour justifier la décomposition de systèmes complexes en sous-systèmes. Ainsi par exemple, dans une longue et savante discussion basée sur des arguments évolutionistes, énergétiques, économiques, historiques, ..., H.A.SIMON [1962] montre que, sauf quelques exceptions "pathologiques", toutes les composantes d'un système ne sont pas en relation, ou si elles le sont, il existe une hiérarchie dans l'intensité de cette relation qui permet leur regroupement en sous-systèmes, sous-sous-systèmes, ..., de façon telle que le comportement à court terme dans chaque sous$^{(k)}$-système est "à peu près" indépendant du comportement à court terme des autres sous$^{(k)}$-systèmes. A long terme, un sous$^{(k)}$-système est influencé par les autres, mais d'une façon agrégée.

De leur côté, W.R.ASHBY et R.C. CONANT justifient la prédominance de la structure hiérarchique : système, sous-système, sous-sous-système, ..., en montrant que c'est justement celle qui offre les meilleures possibilités de manipulation et de traitement de l'information nécessaires pour contrecarrer les perturbations externes [CONA 74], puisque tout autre type de structure demanderait des capacités de calcul et de transmission trop élevées [ASHB 70a, 73]. Enfin, W.R.ASHBY montre dans un petit article que les seuls systèmes complexes qui ont des "chances" d'être stables sont ceux où chaque variable n'interagit directement qu'avec un faible pourcentage des autres (de 10% à 15% maximum !) [ASHB 70b] .

Ce chapitre commence avec l'étude du concept de partition optimale : celle dont le couplage entre les sous-systèmes est minimal. On verra cependant que ce critère d'optimalité mène à une solution triviale, ce qui nous conduit à considérer des optimums locaux. Le reste du chapitre est consacré à la présentation de 5 algorithmes, dont on discute l'optimalité ainsi que le temps et la quantité de mémoire requise pour leur exécution.

Bien entendu, la seule connaissance des variables faisant partir de chaque sous-système ne constitue pas une connaissance complète de la représentation Système-Structure (niveau épistémologique 3,cf.chap.I,§ 2), car il faudrait pour cela déterminer les relations génératrices. Mais la connaissance des "bonnes partitions" du système aide énormément cette tâche.

En effet, la décomposition d'un système en sous-systèmes faiblement couplés permet de concentrer les moyens d'observation, d'analyse et de calcul sur les sous-systèmes, de taille plus réduite, les intéractions étant étudiées dans une phase ultérieure. Elle autorise aussi l'application des méthodes de modélisation par modèles partiels, commande descentralisée, optimisation hiérarchique, ..., dont la convergence est d'autant plus rapide que les sous-systèmes sont faiblement couplés [RICH 75] .

2 - LE CONCEPT DE PARTITION OPTIMALE

2.1. - LA MINIMISATION DE LA TRANSINFORMATION EXTERNE I

R.C. CONANT fut le premier à proposer l'utilisation de la Transinformation Externe $\overleftrightarrow{I}$ comme critère d'optimalité, dans un article qui est devenu une référence obligée dans le domaine [CONA 72]. En effet, si $R = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}(\Sigma)$, $\overleftrightarrow{I}(R)$ mesure la quantité d'information circulant entre les sous-systèmes $R_1, R_2, \ldots, R_r$. Or c'est justement la relation de dépendance ou de causalité entre variables (ou groupes de variables) qui explique principalement cette circulation d'information. Alors, les partitions $M^* \in \mathbb{P}(\Sigma)$ pour lesquelles $\overleftrightarrow{I}(M^*)$ est minimale, ne devraient pas en principe affecter à des sous-systèmes différents les variables fortement liées par des relations de dépendance.

Cela apparait plus clairement dans la thèse de M.RICHETIN [1975, chap.III], qui utilise extensivement une propriété auparavant démontré par W.R.ASHBY [ASHB 69] (c.f. proposition II.13(SV)) :

$$\forall R = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}(\Sigma) :$$

$$\overleftrightarrow{I}(R_1 : R_2 : \ldots : R_r) + \sum_{i=1}^{r} \dot{I}(R_i) = \text{Constante} = \overleftrightarrow{I}(X_1 : X_2 : \ldots : X_N) \qquad [III.1]$$

Ainsi, si l'on minimise le terme $\overleftrightarrow{I}$, cela revient à maximiser le terme $\sum_i \dot{I}$ des Transinformations Internes, donc à choisir des sous-systèmes composés de variables fortement liées.

2.2 - "LOIS GLOBALES" ET "LOIS REGIONALES"

Nous pouvons donner une autre interprétation du critère "Minimiser $\overleftrightarrow{I}$", très informelle peut-être, mais qui aide à mieux comprendre l'importance de cette minimisation.

Comme nous l'avons dit au chapitre II (c.f.§5.), la Transinformation Totale $\overleftrightarrow{I}(X_1:X_2:\ldots:X_N)$ représente en quelque sorte la *"quantité de loi"* qu'on peut espérer formuler sur le système $\Sigma = \{X_1,X_2,\ldots,X_N\}$, et cela indépendamment de la forme spécifique de ces lois (équations de récurrence, équations différentielles, protocoles, relations de causalité, ...). L'affirmation [III.1] nous permet de séparer cette "quantité de loi" en deux parties :

- la première, associée au terme $\overleftrightarrow{I}(R_1:R_2:\ldots:R_r)$, mesure la quantité de loi qui ne parlerait que de la relation entre les sous-systèmes $R_1,R_2,\ldots,R_r$ (*"lois globales"*).

- la deuxième, associée à la somme des termes $\dot{I}(R_i)$, mesure la quantité de loi qui ne parlerait que du comportement à l'intérieur de chacun des sous-systèmes (*"lois régionales"*).

Dans cette optique, l'argumentation qu'on a donné dans l'introduction de ce chapitre pour justifier l'approche de la décomposition, revient à dire que tous les systèmes complexes (sauf quelques exceptions "pathologiques") admettent une décomposition en sous-systèmes telle que la quantité de loi est bien d'avantage locale que globale.

En conclusion donc, l'intérêt de déterminer les partitions $M^* \in \mathbb{P}(\Sigma)$ qui minimisent $\overleftrightarrow{I}$ est que avec ces partitions on peut obtenir une idée assez complète du système (d'autant plus complète que $\overleftrightarrow{I}(M^*)$ est petit) rien qu'en étudiant les "lois régionales" de chaque sous-système, et que ce qui restera à connaitre sera minimal.

2.3 - DECOMPOSITION OPTIMALE TRIVIALE

Cependant, comme le montre Z.ABID [1979,chap.IV], le critère "Minimiser $\overleftrightarrow{I}$" utilisé seul, conduit à une solution triviale. En effet, soit $M^* = \{R_1^*, R_2^*, \ldots, R_r^*\} \in \mathbb{P}(\Sigma)$ une partition optimale. Alors par définition de l'optimalité :

$$\forall\ M = \{R_1, R_2, \ldots, R_K\} \in \mathbb{P}(\Sigma) : \qquad [III.2]$$

$$\overleftrightarrow{I}(M^*) = \overleftrightarrow{I}(R_1^* : R_2^* : \ldots : R_r^*) \leqslant \overleftrightarrow{I}(R_1 : R_2 : \ldots : R_K) = \overleftrightarrow{I}(M).$$

Mais on sait aussi (c.f. prop.II.14 (SV)) que si $M^{\sim} \geqslant M^*$, alors $\overleftrightarrow{I}(M^{\sim}) \leqslant \overleftrightarrow{I}(M^*)$. Cela, avec l'optimalité de M^*, implique que n'importe quelle partition moins fine que M^* est aussi optimale :

(SV) :

$$\forall\ M^{\sim} \geqslant M^* : \left\{ \begin{array}{l} \overleftrightarrow{I}(M^*) \leqslant \overleftrightarrow{I}(M^{\sim}) \text{ par [III.2]} \\ \\ \overleftrightarrow{I}(M^*) \geqslant \overleftrightarrow{I}(M^{\sim}) \text{ par Prop.II.14} \end{array} \right\} \Rightarrow \overleftrightarrow{I}(R^{\sim}) = \overleftrightarrow{I}(R^*).$$

En particulier la partition triviale $M_1 := \{(X_1, X_2, \ldots X_N)\}$ est moins fine que M^*, donc

(SV) : $$\overleftrightarrow{I}(M^*) = \overleftrightarrow{I}(M_1) = 0.$$

Cela veut dire que le critère "Minimiser $\overleftrightarrow{I}$" nous conduit à la solution triviale M_1, et peut être à d'autres "décompositions parfaites", dont les sous-systèmes seraient tout à fait indépendantes. Or les "décompositions parfaites" arrivent très rarement : comme nous l'avons déjà dit, les systèmes complexes admettent en général une décomposition en sous-systèmes faiblement couplés, mais cela ne signifie pas que ce couplage soit pour autant négligeable [SIMO 62, § "Nearly Decomposable Systems"].

2.4 - DECOMPOSITIONS OPTIMALES PAR NIVEAUX

Comme l'a montré le paragraphe ci-dessus, "Minimiser $\overleftrightarrow{I}$" tout seul conduit en général à la solution triviale où la meilleure décomposition est celle qui ne contient qu'un sous-système regroupant toutes les variables. Ceci est d'ailleurs vrai non seulement pour H = Entropie, mais en général pour n'importe quelle Semi-Valuation.

Pour remédier à ce problème, nous proposons la *Minimisation de* $\overleftrightarrow{I}$ *par Niveaux* :

Comme nous l'avons vu au chapitre I, l'ensemble $\mathbb{P}(\Sigma)$ muni de la relation $\leqslant$: "être plus fine que", constitue un treillis. Dans ce treillis on définit le *Niveau* r par l'ensemble

$$\mathbb{P}_r(\Sigma) := \{\text{partitions de } \Sigma \text{ en } r \text{ classes}\},\ r = 1,2,\ldots,N.$$

Il est clair que :

$$\mathbb{P}_1(\Sigma) = \{\{(X_1,X_2,\ldots,X_N)\}\},$$

$$\mathbb{P}_N(\Sigma) = \{\{(X_1),(X_2),\ldots,(X_N)\}\},$$

$$\mathbb{P}_K(\Sigma) = \emptyset \text{ pour } K = N+1,\ N+2,\ldots,$$

$$\mathbb{P}_i(\Sigma) \cap \mathbb{P}_j(\Sigma) = \emptyset \text{ pour } i \neq j,\ 1 \leqslant i,\ j \leqslant N,$$

$$\bigcup_{i=1}^{N} \mathbb{P}_i(\Sigma) = \mathbb{P}(\Sigma).$$

On peut faire une représentation graphique de $\mathbb{P}(\Sigma)$ comme une hiérarchie de niveaux, des arcs entre niveaux voisins correspondant à la relation $\leq$ (figure III.1).

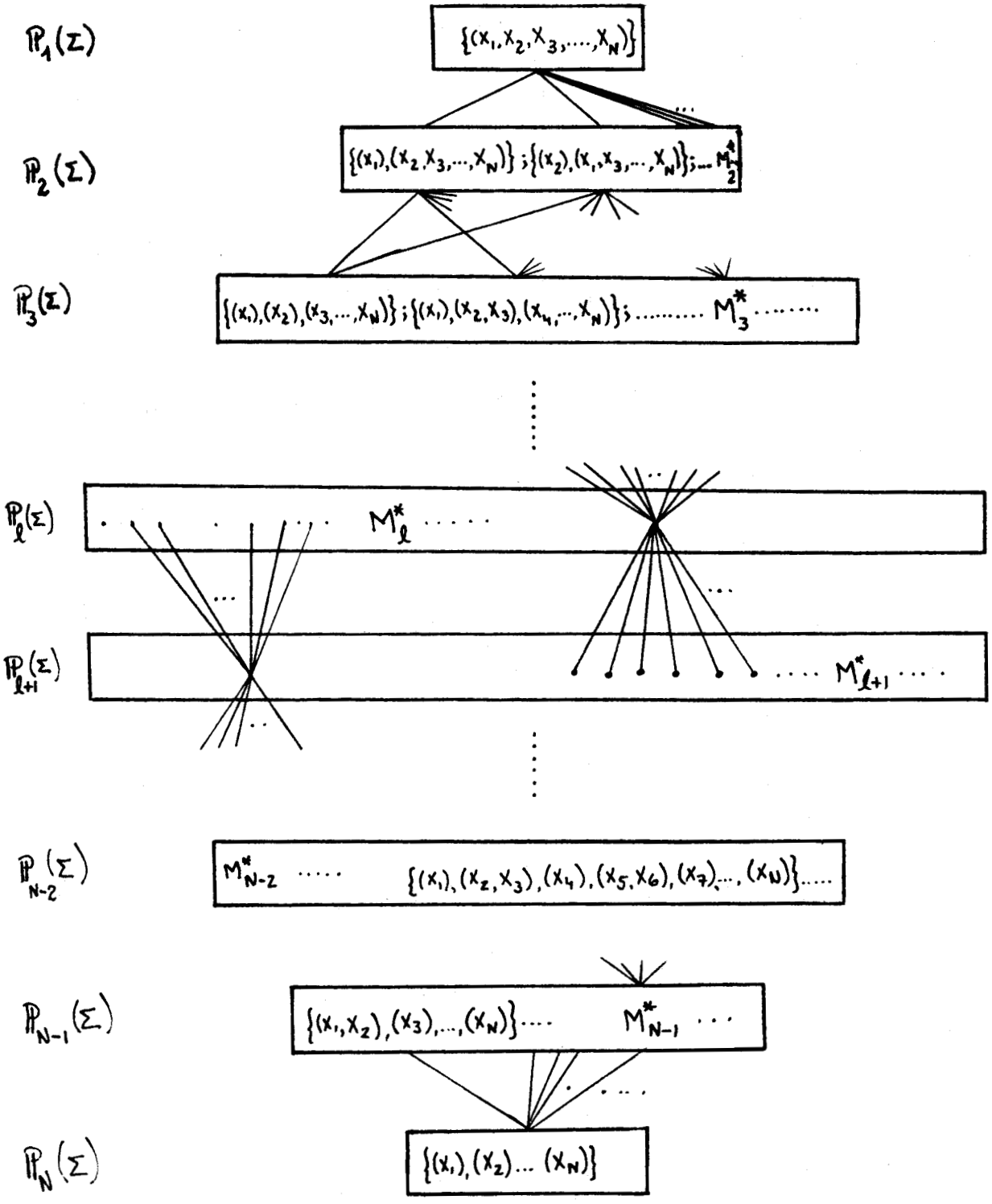


Figure III.1 Le treillis $\mathbb{P}(\Sigma)$

H étant une semi-valuation quelconque définie sur Σ,
le problème de la Décomposition Optimale de Σ est alors énoncé ainsi :

Pour r = 1,2,3,...,N, déterminer la (les) partition(s)
$M_r^* = \{R_1^*, R_2^*, \ldots, R_r^*\} \in \mathbb{P}_r(\Sigma)$, telle que

$$\forall M = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}_r(\Sigma) : \qquad [III.3]$$

$$\overleftrightarrow{I}(M_r^*) = \overleftrightarrow{I}(R_1^*, R_2^*, \ldots, R_r^*) \leq \overleftrightarrow{I}(R_1, R_2, \ldots, R_r) = \overleftrightarrow{I}(M).$$

Les M_r^* sont appellés *Optimums Locaux* . On voit immédiatement que

$$M_1^* = \{(X_1, X_2, \ldots, X_N)\},$$

$$M_N^* = \{(X_1), (X_2), \ldots, (X_N)\},$$

puisque aussi bien $\mathbb{P}_1(\Sigma)$ que $\mathbb{P}_N(\Sigma)$ ne contiennent qu'une seule partition, qui est donc forcément l'optimum local.

Proposition III.1 (SV)

Les optimums locaux $M_1^*, M_2^*, \ldots, M_N^*$ vérifient

$$\overleftrightarrow{I}(M_{r-1}^*) \leq \overleftrightarrow{I}(M_r^*) \qquad , r = 2,3,\ldots,N$$

Démonstration

Soit $M_r^* = \{R_1^*, R_2^*, \ldots, R_r^*\}$. Soit $[M_r^*]' := \{R_1^*, R_2^*, \ldots, R_{iu}^* R_j^*, \ldots, R_r^*\}$, la partition obtenue en regroupant dans une seule deux des classes de M_r^*. Alors $[M_r^*]' \in \mathbb{P}_{r-1}(\Sigma)$, donc

$$\overleftrightarrow{I}(M^*_{r-1}) \leqslant \overleftrightarrow{I}([M^*_r]') \ ;$$

et d'après la proposition II.15 (SV) :

$$\overleftrightarrow{I}(M^*_{r-1}) \leqslant \overleftrightarrow{I}([M^*_r]') = \overleftrightarrow{I}(M^*_r) - \overleftrightarrow{I}(R^*_i:R^*_j) \leqslant \overleftrightarrow{I}(M^*_r)$$

(F.D.)

Remarque
========

Il faut cependant noter que

$$\overleftrightarrow{I}(M^*_{r-1}) \leqslant \overleftrightarrow{I}(M^*_r) \not\Rightarrow M^*_{r-1} \geqslant M^*_r$$

En effet, les optimums locaux ne constituent malheureusement pas une chaîne ordonnée de partitions. Cela rend plus difficile leur détermination, comme on va le voir.

Enfin, notons avant de passer à l'étude des différents algorithmes, que l'approche consistant à minimiser l'intervaluation externe $\overleftrightarrow{I}$ est valable aussi bien pour les systèmes statiques que pour les systèmes dynamiques, c'est-à-dire que dans III.3 sont compris au même titre le cas statique (Σ^E) et le cas dynamique (Σ^D) (c.f.chap.I, §4.2.1 et § 4.2.2). En particulier pour les systèmes dynamiques nous n'aurons nulle pas besoin de l'hypothèse de *déterminisme*.

$$H(\{X_1(t),X_2(t),\ldots,X_N(t)\}/\{X_1(t-1),X_2(t-1),\ldots,X_N(t-1)\})=0$$

qui exige que connaissant les valeurs prises par les variables à un instant donné, il soit toujours possible de déterminer, sans ambiguité la valeur des variables à l'instant suivant. Cette hypothèse de déterminisme est nécessaire aux approches proposées par M.RICHETIN et J.DUFOUR [RICH 75,chap.III,pp.83], [DUFO 79,chap.II, pp.2.38].

3 - DENOMBREMENT DE $\mathbb{P}(\Sigma)$ ET RECHERCHE EXHAUSTIVE

Dans l'ensemble $\mathbb{P}(\Sigma)$ nous ne possédons pas, bien entendu, les structures algébriques nécessaires pour les méthodes classiques d'optimisation (dérivées, gradients, multiplicateurs de Lagrange,...). Il faudra alors que la recherche des optimums locaux se fasse au moyen d'algorithmes qui se déplacent dans le treillis $\mathbb{P}(\Sigma)$. Pour avoir une idée de la complexité de cette tâche, nous estimons le temps qu'un algorithme de Recherche Exhaustive (RE) mettrait pour trouver les optimums locaux M_1^*, M_2^* ,..., M_N^* . [TORO 78]

3.1 - ALGORITHME DE RECHERCHE EXHAUSTIVE (RE)

Imaginons qu'on a stocké dans un fichier F toutes les partitions de Σ. Supposons aussi que l'on a déjà calculé une table avec la valeur de H(S), pour tout $S \in \mathbb{P}^*(\Sigma)$.

L'algorithme de Recherche Exhaustive (RE) lit une à une les partitions contenues dans F,calcule l'Intervaluation Externe $\overleftrightarrow{I}$ de la partition lue, et compare avec la partition qui jusqu'à ce moment était la meilleure. Si la dernière partition lue possède un $\overleftrightarrow{I}$ plus petit, cette partition devient la meilleure partition lue. Tout cela bien sûr, discriminé d'après les niveaux du treillis.

```
. début;
. MEILLEUR_I_i ← + ∞    , i = 1,2,...,N;
. P ← première partition lue du fichier F;
Tant que le fichier F n'est pas fini faire ;
    . i ← nombre de classes de P;
    Si Ī (P) < MEILLEUR_I_i
      alors;
        . MEILLEUR_I_i ← Ī (P);
        . MEILLEURE_PARTITION_i ← P;
    fsi
    . P ← suivante partition lue du fichier F;
ftq;
.Ecrire MEILLEURE_PARTITION_i,MEILLEUR_I_i, i=1,2,...,N;
.fin;
```

3.2 - OPTIMALITE DE L'ALGORITHME RE

Cet algorithme écrit à la fin une liste de N partitions, avec leur intervaluation $\overrightarrow{I}$ respective. Ces partitions seront bien entendu les optimums locaux, puisque l'algorithme aura examiné toutes les partitions possibles (le paragraphe suivant suggère comment construire un tel fichier F, avec toutes les partitions de Σ).

L'optimalité de ce type d'algorithme est bien sûr assurée, quelle que soit la semi-valuation H utilisée pour calculer $\overrightarrow{I}$, et même encore, quel que soit le critère d'optimalité.

3.3 - ORDRE DE L'ALGORITHME RE

Dans les algorithmes que nous présentons dans ce chapitre et dans le suivant, on peut distinguer trois parties :

- Initialisation : où l'on donne des valeurs initiales aux variables, on lit les paramètres et les premières données. Cette partie ne prend que quelques lignes du programme.
- Boucle Principale : c'est le coeur de l'algorithme, où le gros des calculs, comparaisons et décisions sont faits. Elle comporte la lecture de quelques données et/ou l'incrémentation de certains indices, puis les instructions de calcul et de comparaison, et enfin, la décision de sortie qui fait recommencer la boucle, ou bien, mène à la partie résultats.
- Résultats : où l'on écrit les résultats obtenus.

En général, le temps nécessaire pour l'initialisation et l'impression des résultats est pratiquement le même pour les algorithmes correspondants aux différentes méthodes; ceux-ci se distinguent alors par le temps passé dans la Boucle Principale.

A ce propos, les calculs faits dans la boucle principale reviennent en fin de compte à l'examen consécutif d'un certain nombre de "*partitions candidates*", en vue d'en trouver la meilleure. Ce nombre de partitions examinées est appelé *Ordre de l'Algorithme*. Celui-ci constitue un paramètre très important, puisque le temps d'exécution lui sera proportionnel. Plus exactement, le temps d'exécution sera égal à

$$\left[\begin{array}{l}\text{temps requis pour l'examen}\\ \text{d'une "partition candidate"}\end{array}\right] \times \left[\text{ordre de l'algorithme}\right] + \left[\begin{array}{l}\text{temps d'initialisation}\\ \text{et d'impression des}\\ \text{résultats.}\end{array}\right]$$

L'algorithme RE examine toutes les partitions de Σ, avec $|\Sigma| = N$. Calculons donc :

$$|\mathbb{P}_K(\Sigma)| =: S(N,k) :=$$ nombre de partitions en k classes d'un ensemble de N éléments.

Ainsi, du fait que les $\mathbb{P}_k(\Sigma)$ sont disjoints, on a :

$$|\mathbb{P}(\Sigma)| = \sum_{K=1}^{N} |\mathbb{P}_K(\Sigma)| = \sum_{K=1}^{N} S(N,k).$$

Le nombre S(N,k) est appelé nombre de Stirling de deuxième espèce, et il est donné par la relation de récurrence :

$$\begin{cases} S(N,1) = S(N,N) = 1, \text{ quel que soit } N \geq 1, \\ S(N+1,K) = S(N,K-1) + K.S(N,K) \text{ , pour } 2 \leq K \leq N-1, \\ S(N,K) = 0, \text{ pour } K > N. \end{cases} \qquad [III.4]$$

En effet, un ensemble de N éléments possède une seule partition en une classe, et une seule en N classes. De plus, si l'on suppose connues les partitions d'un ensemble de N éléments, les partitions de (N+1) éléments en K classes peuvent être obtenues :

- soit parce que dans chaque partition de N éléments en k-1 classes, on ajoute une nouvelle classe qui ne contient que le $(N+1)^{ème}$ élément. On obtient de cette façon S(N,k-1) partitions.

- soit parce que à chaque partition de N éléments en k classes, on ajoute le $(N+1)^{ème}$ élément à n'importe laquelle de ses k classes. Ainsi, chaque partition de N éléments en k classes peut donner lieu à k partitions de N+1 éléments en k classes, soit au total k.S(N,k).

Comme on peut le voir dans le tableau présenté dans l'Annexe 1, le nombre d'éléments de l'ensemble $\mathbb{P}(\Sigma)$ croit très vite avec $N=|\Sigma|$. Si l'on suppose par exemple qu'un ordinateur peut examiner 1000 partitions par seconde, on pourrait envisager l'utilisation de cet algorithme RE jusqu'au maximum $|\Sigma| = 12$ variables. Dans ce cas $|\mathbb{P}(\Sigma)|= 4'213.597$ partitions, et le temps de calcul serait de

$$4'213.597 \times \frac{1}{1000} \times \frac{1}{60} \times \frac{1}{60} = 1 \text{ heure} + 10 \text{ minutes.}$$

Au delà, le nombre de partitions à examiner, donc le temps de calcul, devient trop grand, voire astronomique. Pour $|\Sigma| = 13$ il y a 27'644.437 partitions, et il faudrait 7 heures + 40 minutes de temps de calcul. Si $|\Sigma| = 20$, le nombre de partitions est de 51"724.158'200.000, qui pourraient être examinées en 1640 ans + 59 jours + 5 heures + 39 minutes!!!

4 - RECHERCHE ASCENDANTE DES OPTIMUMS LOCAUX

La proposition II.15 (sv) dit que si l'on passe d'une partition $R = \{R_1, R_2, \ldots, R_r\}$ du niveau r à une partition du niveau r-1 par regroupement en une seule de deux de ses classes, R_i, R_j, l'intervaluation externe $\overleftrightarrow{I}$ diminue de $\overleftrightarrow{I}(R_i : R_j)$. Cela nous suggère un algorithme qui en partant de la partition triviale $M_N = \{(X_1),(X_2),\ldots(X_N)\}$, traverse les différents niveaux du treillis $\mathbb{P}(\Sigma)$, en essayant à chaque étape de diminuer autant que possible l'intervaluation externe $\overleftrightarrow{I}$ par regroupement des deux sous-systèmes les plus liés [TORO 81].

4.1 - ALGORITHME ASCENDANT : ANALYSE HIERARCHIQUE (AH)

En réalité, le principe de cet algorithme est le même que celui des méthodes de Classification Hiérarchique [CAIL 76, chap.XV] : A partir d'une partition du niveau r, on recherche parmi toutes celles du niveau r-1 moins fines, celle qui minimise le critère d'optimalité. Compte-tenu de la proposition II.15 (sv), cette minimisation est obtenue par la réunion des deux classes R_α et R_β pour lesquelles $\overleftrightarrow{I}(R_\alpha : R_\beta)$ à la valeur la plus élevée.

L'algorithme AH ci-dessous utilise deux vecteurs, M_r et I_r, $r = 1,2,\ldots,N$, où il va mettre la partition trouvée à chaque niveau, ainsi que son intervaluation externe. L'algorithme suppose aussi que l'on connait déjà $H(S)$, $\forall\ S \in \mathcal{P}^*(\Sigma)$.

. début;

. $M_N \leftarrow \{(X_1),(X_2),\ldots,(X_N)\}$;

. $I_N \leftarrow \overleftrightarrow{I}(X_1 : X_2 : \ldots : X_N)$;

. $r \leftarrow N$;

tant que $r \geq 3$ faire ;

. Examiner tous les couples de sous-systèmes de M_r, et trouver les sous-systèmes R_α et R_β, $1 \leq \alpha < \beta \leq r$, qui vérifient :

$$\overleftrightarrow{I}(R_\alpha : R_\beta) \geq \overleftrightarrow{I}(R_i : R_j) \quad \forall\ i,j,\ 1 \leq i < j \leq r ;$$

. $M_{r-1} \leftarrow (M_r \setminus \{R_\alpha, R_\beta\}) \cup \{R_\alpha \cup R_\beta\}$;

. $I_{r-1} \leftarrow I_r - \overleftrightarrow{I}(R_\alpha : R_\beta)$;

. $r \leftarrow r - 1$;

ftq;

. $M_1 \leftarrow \{(X_1, X_2, \ldots, X_N)\}$;

. $I_1 \leftarrow 0$;

. Ecrire M_i, I_i, $i = 1,2,\ldots,N$;

. fin ;

Dans l'annexe 2 on donne une version étendue de cet algorithme où en plus de R_α et R_β, on examine aussi le regroupement d'autres sous-systèmes R_u, R_v tels que $\left| \overleftrightarrow{I}(R_\alpha : R_\beta) - \overleftrightarrow{I}(R_u : R_v) \right| \leq \varepsilon$, pour un $\varepsilon \geq 0$ donné.

4.2 - OPTIMALITE DE L'ALGORITHME AH

L'algorithme AH n'est évidemment pas optimal. On peut seulement être sûr que l'algorithme trouvera les optimums locaux M_N^*, M_1^* (les partitions triviales), ainsi que M_{N-1}^* , puisque dans ces niveaux il aura examiné toutes les partitions. Pour les autres niveaux, N-2, N-3,...,2, on ne peut rien dire puisque l'algorithme n'aura examiné que les

partitions qui lui sont accessibles, c'est-à-dire, celles obtenues par regroupement de deux classes de la partition précédente. Or, comme on l'a remarqué auparavant, les optimums locaux ne constituent pas forcément une chaine ordonnée par la relation $\leqslant$.

4.3 - ORDRE DE L'ALGORITHME AH

Pour un niveau r donné, la boucle principale de l'algorithme AH examine tous les couples de sous-systèmes de la partition M_r, pour trouver les deux sous-systèmes R_α et R_β les plus liés. Cela est équivalent à l'examen de toutes les partitions du niveau r-1 moins fines que M_r, pour trouver celle qui minimise $\overleftrightarrow{I}$.

On peut dire alors que le passage du niveau r au niveau r-1 fait l'examen d'un nombre de partitions égal au nombre de couples de sous-systèmes de la partition précédente, c'est-à-dire

$$\frac{r.(r-1)}{2} \qquad [III.5]$$

Au total donc, le nombre de partitions examinées par l'algorithme AH est :

$$\text{ordre (AH)} := \sum_{r=N}^{2} \frac{r.(r-1)}{2}$$

$$= \frac{1}{2} \sum_{r=1}^{N} (r^2 - r) = \frac{1}{2} \sum_{r=1}^{N} r^2 - \frac{1}{2} \sum_{r=1}^{N} r$$

$$= \frac{1}{2} \cdot \left(\frac{N.(N+1).(2N+1)}{6}\right) - \frac{1}{2} \cdot \left(\frac{N.(N+1)}{2}\right)$$

$$= \frac{N^3 - N}{6} \qquad [III.6]$$

L'annexe 1 montre les valeurs de l'ordre (AH), pour $|\Sigma| = N \leqslant 20$. On voit que le nombre de partitions ne croit pas trop vite, et si l'on suppose à nouveau que l'ordinateur prend un millième

de seconde dans l'examen de chaque partition, on pourrait envisager l'utilisation de l'algorithme AH pour $|\Sigma| = N \leqslant 378$ variables. En effet, pour $|\Sigma| = 378$, ordre (AH) = 9'001.629, et le temps de calcul nécessaire serait 2 heures + 30 minutes + 1 seconde.

5 - RECHERCHE DESCENDANTE DES OPTIMUMS LOCAUX

La méthode que nous présentons maintenant est en quelque sorte l'inverse de celle décrite dans le paragraphe 4. Il s'agit d'un algorithme de segmentation, procédant par fractionnements successifs des classes [ABID 79, chap.IV; STAR 81; TORO 79,81]

5.1 - ALGORITHME DESCENDANT : DIVISION RECURSIVE (DR)

La proposition II.16 (sv) dit que si l'on passe d'une partition $M_r = \{ R_1, R_2, \ldots, R_r \}$ du niveau r à une partition du niveau r+1 par l'éclatement d'une de ses classes en deux, l'intervaluation externe $\overleftrightarrow{I}$ augmente de l'intervaluation entre les deux classes créées. En partant de la partition triviale $M_1 = \{ (X_1, X_2, \ldots, X_N) \}$, l'idée est alors de traverser les différents niveaux du treillis en essayant à chaque étape de créer les deux sous-systèmes qui fassent augmenter l'intervaluation externe le moins possible.

L'algorithme DR ci-dessous stocke dans les vecteurs M_r et I_r, $i = 1,2,\ldots,N$ la partition qu'il trouve à chaque niveau, ainsi que son intervaluation externe. L'algorithme suppose aussi que l'on connait déjà $H(S)$, $\forall\, S \in \mathcal{P}^*(\Sigma)$

. début;

. $M_1 \leftarrow \{(X_1, X_2, \ldots, X_N)\}$;

. $I_1 \leftarrow 0$;

. $r \leftarrow 1$;

tant que $r \leqslant N-2$ faire;

- . pour chaque $R_i \in M_r$ examiner tous les $S \subseteq R_i, \emptyset \neq S \neq R_i$, et trouver le R_α et le $S^* \subseteq R_\alpha$, $\emptyset \neq S^* \neq R_\alpha$, tels que
  $\overleftrightarrow{I}(R_\alpha \setminus S^* : S^*) \leqslant \overleftrightarrow{I}(R_i \setminus S : S)$ $\forall\, R_i \in M_r$, $\forall\, S \subseteq R_i, \emptyset \neq S \neq R_i$;
- . $M_{r+1} \leftarrow (M_r \setminus \{R_\alpha\}) \cup \{R_\alpha \setminus S^*, S^*\}$;
- . $I_{r+1} \leftarrow I_r + \overleftrightarrow{I}(R_\alpha \setminus S^* : S^*)$;
- . $r \leftarrow r+1$;

ftq ;

. $M_N \leftarrow \{(X_1),(X_2),\ldots,(X_N)\}$;

. $I_N \leftarrow \overleftrightarrow{I}(X_1 : X_2 : \ldots : X_N)$;

. Ecrire M_r, I_r, pour r = 1,2,..., N;

. fin;

L'annexe 3 détaille cet algorithme et utilise certaines tables, pour éviter la répétition des calculs implicites dans la première instruction de la boucle principale ("Pour chaque $R_i \in M_r$,...").

5.2 - OPTIMALITE DE L'ALGORITHME DR

Aussi bien que AH, l'algorithme DR n'est pas optimal. Les seuls optimums locaux trouvés par DR sont M_1^* et M_N^* (les partitions triviales), ainsi que M_2^* , puisque dans ces niveaux il aura examiné toutes les partitions. Pour les autres niveaux, 3,4,... N-1, l'algorithme n'aura examiné que les partitions qui lui sont accessibles, c'est-à-dire, celles obtenues par l'éclatement en deux d'un des sous-systèmes de la partition précédente.

Toutefois, cet algorithme DR a des "bonnes chances" de trouver les partitions optimales qui ont une petite intervaluation externe. Sans que ceci prétende être une démonstration, imaginons qu'il existe une partition optimale M_r^* au niveau r pour laquelle $\overleftrightarrow{I}(M_r^*) = \varepsilon$ est "particulièrement"petite. Alors toute partition moins fine que M_r^* aura une intervaluation externe $\overleftrightarrow{I}$ plus petite ou égale à ε (c.f. Proposition II.14(sv)).

Ces partitions moins fines que M_r^* forment des chaînes entre M_1^* et M_r^* qui traversent tous les niveaux (fig.III.2)

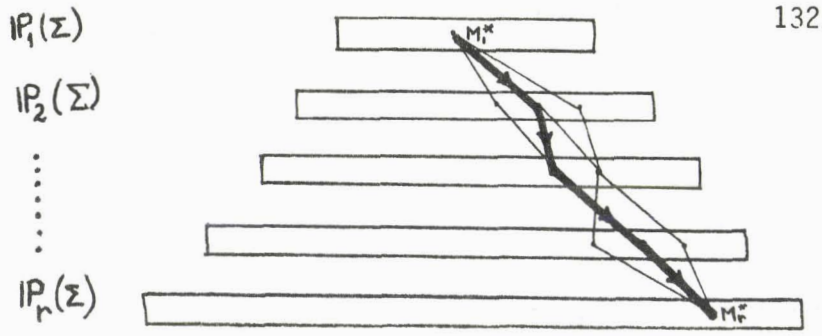


Figure III.2 Descente vers une partition optimale

entre M_1^* et M_r^* . Alors, sous "l'hypothèse" d'un ε "particulièrement" petit,on pourrait espérer que l'algorithme descende par une de ces chaînes jusqu'à M_r^* , puisque dès le niveau $\mathbb{P}_2(\Sigma)$ et à chacun des niveaux suivants il trouvera des partitions dont l'intervaluation est inférieure ou égale à ε.

Ces "chances accrues" d'optimalité se reflètent dans un temps de calcul bien plus élevé que celui de l'algorithme AH.

5.3 - ORDRE DE L'ALGORITHME DR

A chaque niveau l'algorithme DR examine toutes les décompositions en deux de chacun des sous-systèmes. Cela est équivalent à examiner toutes les partitions obtenues par l'éclatement en deux d'un des sous-systèmes de la partition précédente.

Mais comme le montre l'annexe 3, à partir du niveau $\mathbb{P}_2(\Sigma)$, il suffit d'examiner uniquement les partitions en deux des deux sous-systèmes créés dans l'étape antérieure. Soient R_1 et R_2 ces sous-systèmes. Le nombre d'alternatives examinées est alors

$$f(|R_1|,|R_2|) := \frac{2^{|R_1|} - 2}{2} + \frac{2^{|R_2|} - 2}{2}$$

En effet, pour R_1 il faut tester toutes les partitions de la forme $\{R_1 \setminus S, S\}$, avec $S \subseteq R$, $\emptyset \neq S \neq R_1$. Il y a $2^{|R1|}$ sous-ensembles de R_1, dont on élimine $S = \emptyset$ et $S = R_1$; la division par 2 est dûe au fait que $\overleftrightarrow{I}(R_1 \setminus S : S) = \overleftrightarrow{I}(S : R_1 \setminus S)$. Pour R_2 le raisonnement est le même.

Il n'est pas possible de prédire les cardinaux de R_1 et R_2. On peut quand même dire que, au niveau r,

$$|R_1| + |R_2| \leqslant N - (r-2)$$

puisqu'il faut que chacun des autres r-2 sous-systèmes aient au moins une variable. Alors, vu la nature exponentielle de f, on peut dire

$$f(|R_1|,|R_2|) \leqslant f(N - (r-2) - 1, 1) = 2^{N-r} - 1 \qquad [III.7]$$

Ainsi, le nombre de partitions examinées pour passer du niveau r ($\geqslant 2$) au niveau r+1 est au maximum $2^{N-r} - 1$. Par ailleurs, pour passer de $\mathbb{P}_1(\Sigma)$ à $\mathbb{P}_2(\Sigma)$ il faut tester $(2^N-2)/2 = 2^{N-1} - 1$ partitions. En faisant la somme de r=1 à r=N-1 on obtient alors

$$\text{Ordre (DR)} \leqslant (2^{N-1}-1) + \sum_{r=2}^{N-1} (2^{N-r}-1) = \sum_{r=1}^{N-1} (2^{N-r}-1)$$

$$= \sum_{r=1}^{N-1} 2^{N-r} - (N-1) = \sum_{j=1}^{N-1} 2^j - (N-1)$$

$$= \frac{2^N - 2}{2 - 1} - (N-1) = 2^N - 2 - N + 1$$

$$= 2^N - N - 1 \qquad [III.8]$$

Il faut remarquer que ce résultat ne constitue qu'une bonne supérieure de l'ordre (DR), correspondant au pire des cas. Toutefois, si l'on suppose un temps d'un millième de seconde pour l'examen de chaque partition, on pourrait utiliser l'algorithme DR pour $|\Sigma| \leqslant 23$. En effet pour $|\Sigma| = 23$ il y aurait au maximum 8'388.538 partitions à examiner, tâche qui necessiterait au plus 2 heures + 19 minutes + 49 secondes.

6 - APPROXIMATIONS SUCCESSIVES DES OPTIMUMS LOCAUX

Les algorithmes AH et DR qui viennent d'être présentés peuvent se résumer respectivement à une application répétée des deux opérateurs suivants :

* OPERATEUR AH [III.9]

$$AH : \mathbb{P}_r(\Sigma) \longrightarrow \mathbb{P}_{r-1}(\Sigma), \quad 2 \leqslant r \leqslant N$$

$$R=\{R_1,\ldots,R_r\} \mapsto AH(R):=\{R_1,..,R_\alpha \cup R_\beta,..,R_r\}$$

$$=(R\setminus\{R_\alpha,R_\beta\})\cup\{R_\alpha \cup R_\beta\},$$

avec $R_\alpha, R_\beta \in R$ tels que :

$$\overleftrightarrow{I}(R_\alpha : R_\beta) = \operatorname*{Max}_{i,j} \overleftrightarrow{I}(R_i:R_j) \qquad 1 \leqslant i < j \leqslant r$$

D'après la proposition II.15 (sv) on a

$$\overleftrightarrow{I}(AH(R)) = \overleftrightarrow{I}(R) - \overleftrightarrow{I}(R_\alpha : R_\beta) \qquad [III.10]$$

* OPERATEUR DR [III.11]

$$DR : \mathbb{P}_r(\Sigma) \longrightarrow \mathbb{P}_{r+1}(\Sigma), \quad 1 \leqslant r \leqslant N-1$$

$$R=\{R_1,\ldots,R_r\} \mapsto DR(R)=\{R_1,..,R_\alpha \setminus S^*,S^*,..,R_r\}$$

$$=(R\setminus\{R_\alpha\})\cup\{R_\alpha\setminus S^*,S^*\}$$

avec $R_\alpha \in R$, $S^* \subsetneq R_\alpha$, $\emptyset \neq S^* \neq R_\alpha$, et tels que :

$$\overleftrightarrow{I}(R_\alpha\setminus S^* : S^*) \leqslant \overrightarrow{I}(R_i\setminus S : S) \;\forall\, R_i \in R, \forall\, S \subsetneq R_i, \emptyset \neq S \neq R_i.$$

D'après la proposition II.16 (sv) on a

$$\overleftrightarrow{I}(DR(R)) = \overleftrightarrow{I}(R) + \overleftrightarrow{I}(R_{\alpha} \backslash S^* : S^*) \qquad [III.12]$$

6.1 - OPERATEURS CONTRACTANTS

La composition des opérateurs AH et DR nous permet de définir les deux nouveaux opérateurs :

$$AH \circ DR : \mathbb{P}_r(\Sigma) \longrightarrow \mathbb{P}_r(\Sigma) \qquad [III.13]$$
$$R \longmapsto AH \circ DR(R) := AH(DR(R))$$

$$DR \circ AH : \mathbb{P}_r(\Sigma) \longrightarrow \mathbb{P}_r(\Sigma) \qquad [III.14]$$
$$R \longmapsto DR \circ AH(R) := DR(AH(R))$$

pour $2 \leqslant r \leqslant N-1$

Proposition III.2 (SV)

Soit Q une partition quelconque du niveau r, $2 \leqslant r \leqslant N-1$. Alors

① $\overleftrightarrow{I}(AH \circ DR(Q)) \leqslant \overleftrightarrow{I}(Q)$,

② $\overleftrightarrow{I}(DR \circ AH(Q)) \leqslant \overleftrightarrow{I}(Q)$.

Démonstration :

① Soit $Q = \{Q_1, Q_2, \ldots, Q_r\}$. Moyennant une rénumérotation éventuelle, on peut toujours écrire :

$$AH(Q) = \{Q_1 \cup Q_2, Q_3, \ldots, Q_r\},$$

avec

$$\overleftrightarrow{I}(Q_1 : Q_2) \geq \overleftrightarrow{I}(Q_i : Q_j), \quad \forall\, Q_i, Q_j \in Q,$$
$$\overleftrightarrow{I}(AH(Q)) = \overleftrightarrow{I}(Q) - \overleftrightarrow{I}(Q_1 : Q_2).$$

Calculons à présent $DR(AH(Q)) = DR(\{Q_1 \cup Q_2, Q_3, \ldots, Q_r\})$.

Deux cas peuvent se présenter :

1er cas : L'application de DR divise le premier sous-système $Q_1 \cup Q_2$. On a alors :

$$DR(AH(Q)) = \{(Q_1 \cup Q_2) \setminus S^*, S^*, Q_3, \ldots, Q_r\}$$

avec $S^* \subseteq Q_1 \cup Q_2$, $\emptyset \neq S^* \neq Q_1 \cup Q_2$, et

$$\overleftrightarrow{I}((Q_1 \cup Q_2) \setminus S^* : S^*) \leq \begin{cases} \overleftrightarrow{I}((Q_1 \cup Q_2) \setminus S : S), \ \forall S \subseteq Q_1 \cup Q_2, \ \emptyset \neq S \neq Q_1 \cup Q_2, \\ \overleftrightarrow{I}(Q_i \setminus S : S), \ \forall Q_i, \ 3 \leq i \leq r, \ \forall S \subseteq Q_i, \ \emptyset \neq S \neq Q_i. \end{cases}$$

En particulier on a :

$$\overleftrightarrow{I}((Q_1 \cup Q_2) \setminus S^* : S^*) \leq \overleftrightarrow{I}((Q_1 \cup Q_2) \setminus Q_2 : Q_2) = \overleftrightarrow{I}(Q_1 : Q_2)$$

ce qui nous permet de dire :

$$\overleftrightarrow{I}(DR(AH(Q))) = \overleftrightarrow{I}(AH(Q)) + \overleftrightarrow{I}((Q_1 \cup Q_2) \setminus S^* : S^*)$$
$$\leq \overleftrightarrow{I}(AH(Q)) + \overleftrightarrow{I}(Q_1 : Q_2) = \overleftrightarrow{I}(Q)$$

2ème cas : L'application de DR divise un autre sous-système.

Soit Q_3 celui-ci. On a :

$$DR(AH(Q)) = \{Q_1 \cup Q_2, Q_3 \setminus S^*, S^*, Q_4, \ldots, Q_r\}$$

avec $S^* \subsetneq Q_3$, $\emptyset \neq S^* \neq Q_3$

$$\overleftrightarrow{I}(Q_3 \setminus S^* : S^*) \leqslant \begin{cases} \overleftrightarrow{I}(Q_3 \setminus S : S), \ \forall\, S \subsetneq Q_3, \ \emptyset \neq S \neq Q_3, \\ \overleftrightarrow{I}(Q_i \setminus S : S), \ \forall Q_i, \ 4 \leqslant i \leqslant r, \ \forall\, S \subsetneq Q_i, \ \emptyset \neq S \neq Q_i, \\ \overleftrightarrow{I}((Q_1 \cup Q_2) \setminus S : S), \ \forall\, S \subsetneq Q_1 \cup Q_2, \ \emptyset \neq S \neq Q_1 \cup Q_2 \end{cases}$$

En particulier, on a :

$$\overleftrightarrow{I}(Q_3 \setminus S^* : S^*) \leqslant \overleftrightarrow{I}((Q_1 \cup Q_2) \setminus Q_2 : Q_2) = \overleftrightarrow{I}(Q_1 : Q_2)$$

ce qui nous permet de dire que

$$\begin{aligned} \overleftrightarrow{I}(DR(AH(Q))) &= \overleftrightarrow{I}(AH(Q)) + \overleftrightarrow{I}(Q_3 \setminus S^* : S^*) \\ &\leqslant \overleftrightarrow{I}(AH(Q)) + \overleftrightarrow{I}(Q_1 : Q_2) = \overleftrightarrow{I}(Q). \end{aligned}$$

La proposition est ainsi démontrée pour DRoAH. Pour l'affirmation (2), concernant AHoDR, la démonstration est en tout point calquée sur celle-ci.

(F.D.)

Proposition III.3 (SV)

Toute M_r^*, partition optimale du niveau r, est un point fixe des applications DRoAH et AHoDR, c'est-à-dire qu'on a :

$$\overleftrightarrow{I}(DRoAH(M_r^*)) = \overleftrightarrow{I}(AHoDR(M_r^*)) = \overleftrightarrow{I}(M_r^*), \ \forall_r \quad 2 \leqslant r \leqslant N-1.$$

Démonstration :

DRoAH (M_r^*) et AHoDR(M_r^*) sont deux partitions appartenant à $\mathbb{P}_r(\Sigma)$; leur intervaluation externe $\overleftrightarrow{I}$ est alors supérieure ou égale à celle de la partition optimale M_r^*. Par ailleurs, l'utilisation de la proposition précédente fournit l'inégalité dans l'autre sens, ce qui prouve l'égalité annoncée.

(F.D.)

REMARQUE : Parmi tous les opérateurs de la forme $AH^K oDR^K$ ou $DR^K oAH^K$, seul celui correspondant à K=1 possède les propriétés précédentes.

6.2 - ALGORITHME D'APPROXIMATIONS SUCCESSIVES (AS)

On dira qu'une partition $M_r \in IP_r(\Sigma)$ est une *Partition stable* ssi c'est un point fixe à la fois pour les opérateurs DRoAH et AHoDR.

On vient de voir que toute partition ou décomposition optimale M_r^* $(2 \leqslant r \leqslant N-1)$ est une décomposition stable; mais il faut remarquer qu'il ne s'agit là que d'une condition nécessaire d'optimalité: il se peut qu'il y ait des décompositions stables qui ne soient pas des optimums locaux.

L'algorithme d'Approximations successives que nous proposons recherche, pour chacun des niveaux du treillis $IP(\Sigma)$, une partition vérifiant la condition nécessaire d'optimalité (partition stable). Pour cela, dans chaque niveau on se donne une décomposition initiale quelconque, et on applique les opérateurs AHoDR et/ou DRoAH autant de fois que cela est possible jusqu'à la stabilisation. La convergence est assurée par le caractère contractant des opérateurs appliqués.

Plus concrètement, les boucles principales des algorithmes AH et DR deviennent des sous-programmes, appelés par l'algorithme AS :

```
. début;
. r ← 2;
tant que r ≤ N-1 faire;
      . Générer (ou lire) une partition M ε IP_r(Σ);
      . Stable ← "faux";
      tant que ¬ Stable faire;
          tant que AHoDR (M) ≠ M faire;
                . M ← AHoDR (M);
          ftq;
          tant que DRoAH (M) ≠ M faire;
                . M ← DRoAH (M);
          ftq;
          Si AHoDR (M) = DRoAH (M) = M
             alors;
                  . Stable ← "vrai";
          fsi;
      ftq;
      . Ecrire M,Ĩ(M);
      . r ← r+1;
ftq;
. fin;
```

La démarche de cet algorithme AS est illustrée par la figure III.3 :

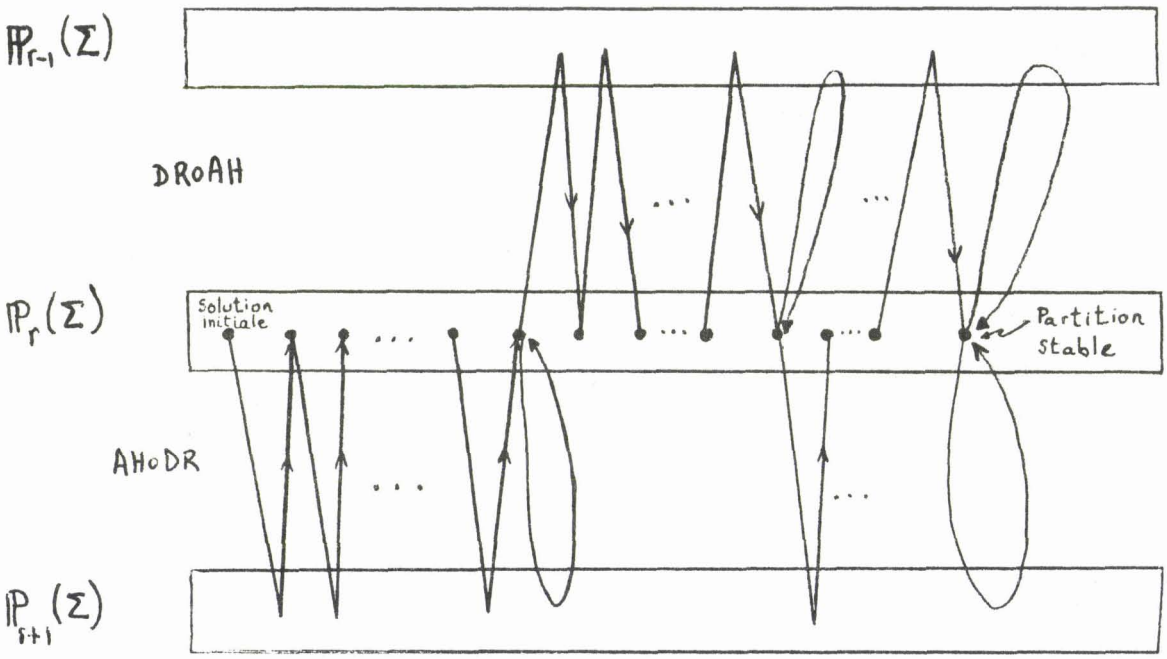


Figure III.3 Démarche de l'algorithme AS

Comme pour les méthodes du type "*Nuées Dynamiques*" [DIDA 79], dans chaque niveau, les calculs pourraient être répétés pour plusieurs partitions initiales tirées au hasard. On obtiendrait ainsi plusieurs partitions stables. Celles-ci permettraient de déterminer les *formes fortes*, c'est-à-dire, les sous-ensembles de variables qui apparaissent toujours dans les mêmes sous-systèmes. Enfin, si une seule partition doit être retenue, on choisira naturellement la partition stable qui possède la plus petite intervaluation $\overleftrightarrow{I}$.

6.3 - OPTIMALITE DE L'ALGORITHME AS

L'algorithme AS fournit pour chaque niveau une (ou plusieurs) partitions stables. Parmi celles-ci, nous pouvons être sûr que pour les niveaux 1 et N (triviallement), ainsi que pour les niveaux 2 et N-1,

les optimums locaux seront trouvés, étant donné que AS réunit les propriétés de DR et AH.

Pour les autres niveaux r, $3 \leqslant r \leqslant N-2$, les partitions trouvées possèdent une propriété nécessaire d'optimalité. La question qu'on se pose alors est de savoir si l'ensemble des partitions stables est "petit", ou si par contre, presque toutes les partitions sont stables. C'est une question ouverte. Nous pouvons dire seulement que dans plusieurs exemples que nous avons développé, il a suffit de déterminer à peine quelques partitions stables pour trouver l'optimum local. Par ailleurs, l'intervaluation externe des partitions stables était en général très près de la valeur optimale.

6.4 - ORDRE DE L'ALGORITHME AS

Les quelques exemples que nous avons étudié nous ont montré aussi que lors du processus de stabilisation à un niveau r, $3 \leqslant r \leqslant N-2$, on est amené à appliquer AHoDR ou DRoAH dans un ordre assez irrégulier : on applique AHoDR jusqu'à l'obtention d'une partition fixe pour cet opérateur, ensuite on applique DRoAH jusqu'à obtenir à nouveau une partition fixe pour celui-ci, à nouveau AHoDR, puis DRoAH,..., jusque stabilisation pour les deux opérateurs. On suppose donc que en moyenne β applications d'un opérateur de contraction reviennent à $\beta/2$ application de AHoDR et $\beta/2$ de DRoAH. Pour les niveaux r = 2 et r = N-1, une seule application de DR et de AH, respectivement, suffisent pour trouver la partition optimale du niveau.

Cela dit, soit $A(N,r,\beta/2)$ (respect.$B(N,r,\beta/2)$) la plus petite borne supérieure du nombre de partitions examinées par $\beta/2$ applications de l'opérateur AHoDR (respect.DRoAH) au niveau r du treillis de partitions de N variables ($3 \leqslant r \leqslant N-2$). En combinant les résultats [III.5] et [III.7] on peut dire que :

$$A(N,r,\beta/2) = \frac{\beta}{2}\left(\underbrace{(2^{N-r}-1)}_{\text{DR:du niveau r au niveau r+1}} + \underbrace{\frac{(r+1).r}{2}}_{\text{AH:du niveau r+1 au niveau r}} \right) \qquad [III.15]$$

$$B(N,r,\beta/2) = \frac{\beta}{2} \cdot \left(\underbrace{\frac{r.(r-1)}{2}}_{\text{AH:du niveau r au niveau r+1}} + \underbrace{(2^{N-(r-1)}-1)}_{\text{DR:du niveau r+1 au niveau r}} \right) \qquad [III.16]$$

$$= A(N,r-1,\beta/2).$$

Enfin, supposons qu'on se donne α partitions initiales à chaque niveau r, $3 \leq r \leq N-2$, et qu'à chacune d'elles on applique β fois les opérateurs de contraction. Le nombre total de partitions examinées, autrement dit, l'ordre de AS, est borné supérieurement par :

$$\text{ordre (AS)} \leq \qquad [III.17]$$

$$\underbrace{\frac{2^N-2}{2}}_{\text{niveau r=2}} + \underbrace{\frac{N.(N-1)}{2}}_{\text{niveau r=N\_1}} + \underbrace{\sum_{r=3}^{N-2} \alpha.\left(A(N,r,\beta/2) + B(N,r,\beta/2)\right)}_{\text{niveau 3 à N-2 (si N} \geq 5)}$$

L'Annexe 1 présente le calcul de Ordre (AS) pour $N \leq 20$, en supposant $\alpha.\beta = 18$ (qui correspondrait, disons, à 3 partitions initiales par niveau, auxquelles on appliquerait 6 fois un opérateur de contraction). Si l'on suppose à nouveau un temps d'un millième de seconde pour l'examen de chaque partition candidate, l'algorithme AS pourrait être utilisé jusqu'à $|\Sigma| = N = 20$: dans ce cas le nombre de partitions à examiner est au maximum de 7'620.905, tâche qui nécessite 2h + 7min + 15sec. de calcul.

7 - UN EXEMPLE D'APPLICATION DES ALGORITHMES AH, DR, ET AS

La présentation de cet exemple fait appel à l'analogie qu'on a établie entre les Semi-Valuations Monotones et la Théorie de la Mesure (Paragraphe 9 du chap.II). Il s'agit d'un système de 7 variables, dont la représentation graphique est donnée ci-dessous.

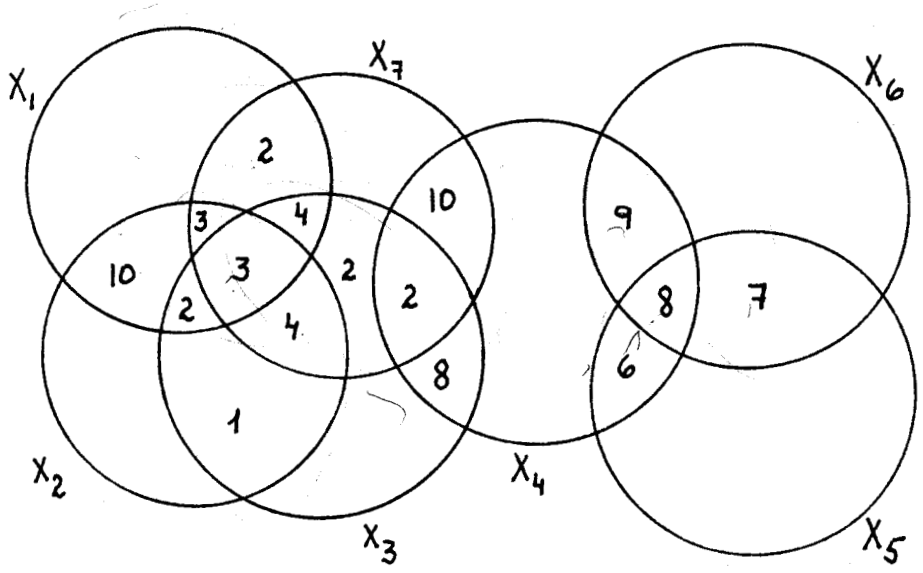


Figure III.4 Représentation graphique d'un système de 7 variables

Les chiffres indiquent la "surface" ou entropie contenue dans chaque secteur élémentaire. Ainsi par exemple

$\overleftrightarrow{I}(X_1:X_2) = 10 + 2 + 3 + 3 = 18,$

$\overleftrightarrow{I}(\{X_1,X_2\} : X_3) = 1 + 2 + 3 + 4 + 4 = 14,$

$\overleftrightarrow{I}(X_4:X_5:X_6) = 9 + 6 + 7 + 2x8 = 38,$

$\overleftrightarrow{I}(\{X_1,X_3\} : \{X_5,X_6\}) = 0,$

$\overleftrightarrow{I}(X_1:X_2:X_3:X_4:X_5:X_6:X_7) = 10 + 2 + 1 + 2 + 10 + 8 + 9 + 6 + 7$
$+ 2x3 + 2x2 + 2x4 + 2x4 + 2x2 + 2x8$
$+ 3x3 = 110.$

L'application des algorithmes AH et DR à ce système fournit les résultats que l'on résume dans les *dendogrammes* [CAIL 76, chap.15] suivants :

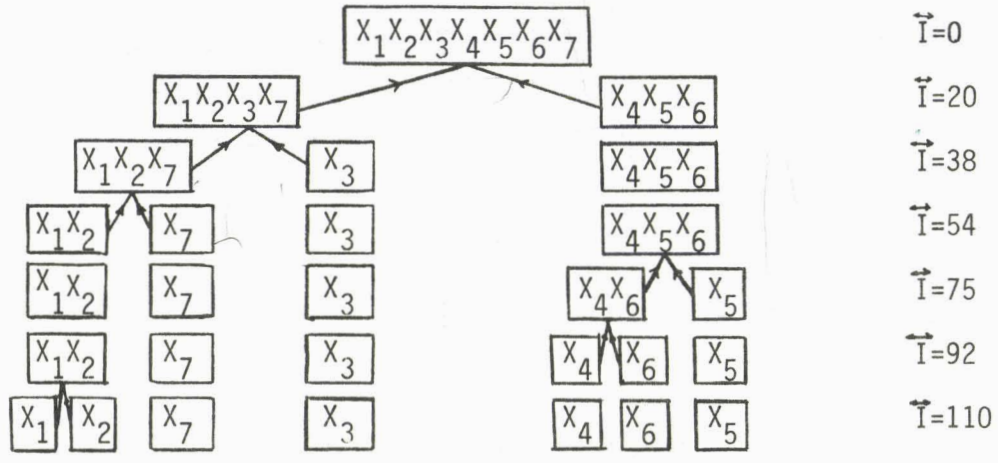


Figure III.5 Résultats de l'algorithme AH

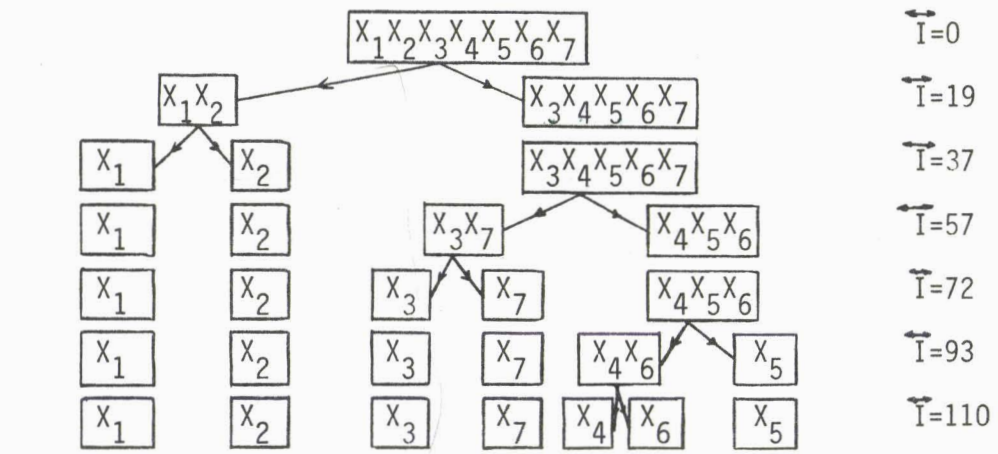
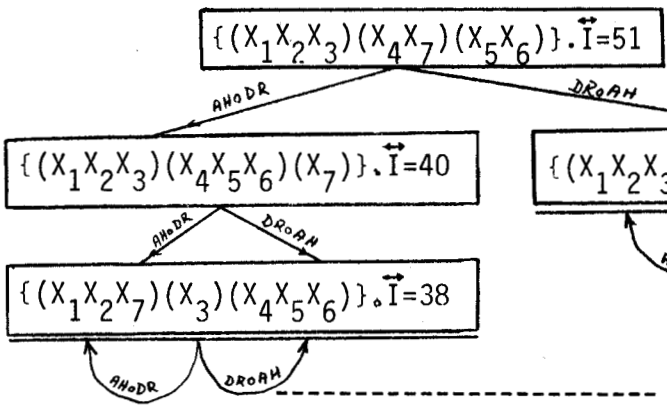


Figure III.6 Résultats de l'algorithme DR

Comme on le voit, pour un niveau donné r, $3 \leqslant r \leqslant N-2$, tantôt DR, tantôt AH peuvent fournir une meilleure partition (pour r=2 c'est DR qui trouve l'optimum, et pour r=N-1 c'est AH qui le trouve). En fait, dans plusieurs exemples que nous avons calculés aucune "superiorité" d'un de ces algorithmes par rapport à l'autre ne s'est dégagée, et cela malgré le fait que DR examine bien plus de partitions candidates que AH. En général, ces deux algorithmes trouvent des bonnes partitions, dont l'intervaluation externe $\overleftrightarrow{I}$ est assez près de celle des optimums locaux.

Voyons à présent l'application répétée des opérateurs contractants AHoDR et DRoAH sur deux partitions du niveau 3 choisies au hasard :

1ère solution de départ:



2èmè solution de départ:

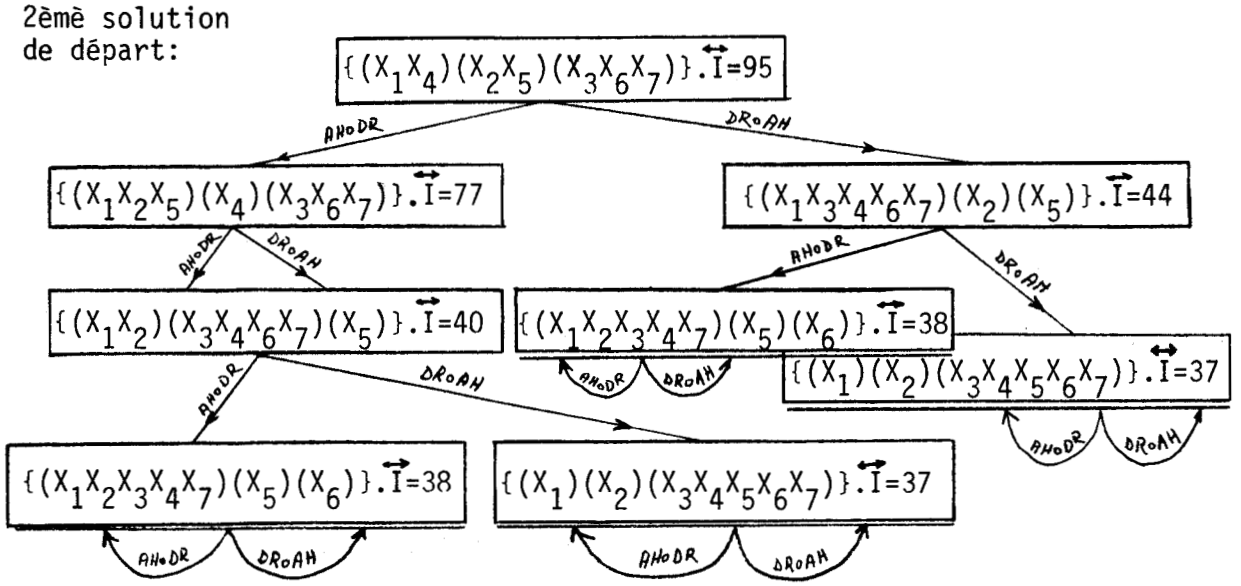


Figure III.7 Application des Opérateurs Contractants.

Cet exemple montre une partition stable, $\{(X_1X_2X_3X_4X_7)(X_5)(X_6)\}$, qui n'est pas un optimum local.

Aussi bien dans cet exemple que dans d'autres un peu plus compliqués que nous avons étudié, on voit qu'en général la convergence est atteinte en un petit nombre de pas, même pour des partitions initiales assez mauvaises. Nous avons trouvé aussi que généralement les partitions stables possèdent une intervaluation externe $\overleftrightarrow{I}$ très proche de celle des optimums locaux.

8 - RECHERCHE DES OPTIMUMS LOCAUX PAR PROGRAMMATION DYNAMIQUE

Une des propriétés montrées au chapitre II va nous permettre d'exprimer l'intervaluation externe $\overleftrightarrow{I}$ sous une forme additive qui autorise l'application de la méthode de Programmation Dynamique. Ce paragraphe est inspirée d'un approche proposée par [STAR 79, chap.1].

8.1 - LA FONCTION "DECOMPOSITION OPTIMALE"

Nous allons étudier la fonction $\mathfrak{D}_r$, qui à chaque sous-ensemble S de variables de Σ fait correspondre la partition optimale de S en r classes. Si l'on appelle $\mathbb{P}_r(S)$ l'ensemble de partitions de S en r classes, la définition formelle de $\mathfrak{D}_r$, r=1,2,...,N, est la suivante :

$$\mathfrak{D}_r : \mathcal{P}(\Sigma) \longrightarrow \mathcal{P}(\Sigma)^r := \underbrace{\mathcal{P}(\Sigma) \times \ldots \times \mathcal{P}(\Sigma)}_{\text{r-fois}}$$

$$S \longmapsto \mathfrak{D}_r(S)$$

où

$$\mathfrak{D}_r(S) := \begin{cases} \text{Si } |S| \geqslant r \text{ alors :} \\ \quad - \mathfrak{D}_r(S) \in \mathbb{P}_r(S), \text{ et} \\ \quad - \forall R = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}_r(S) : \overleftrightarrow{I}(\mathfrak{D}_r(S)) \leqslant \overleftrightarrow{I}(R). \\ \text{Si } |S| < r \text{ alors :} \\ \quad - \mathfrak{D}_r(S) \text{ est "indéfini", et} \\ \quad - \overleftrightarrow{I}(\mathfrak{D}_r(S)) := + \infty. \end{cases}$$

On voit immédiatement que

$$\forall S \neq \emptyset : \mathfrak{D}_1(S) = \{(S)\}, \ \overleftrightarrow{I}(\mathfrak{D}_1(S)) = 0;$$

$$\forall S \neq \Sigma : \mathfrak{D}_N(S) = \text{"indéfini"}, \ \overleftrightarrow{I}(\mathfrak{D}_N(S)) = + \infty;$$

$$\mathfrak{D}_N(\Sigma) = \{(X_1),(X_2),\ldots,(X_N)\}, \ \overleftrightarrow{I}(\mathfrak{D}_N(\Sigma)) = \overleftrightarrow{I}(X_1:\ldots:X_N).$$

L'importance de cette fonction $\mathfrak{D}_r$ réside dans le fait qu'elle permettrait d'obtenir les optimums locaux :

$$\mathfrak{D}_1(\Sigma) = M_1^*, \mathfrak{D}_2(\Sigma) = M_2^*, \ldots, \mathfrak{D}_{N-1}(\Sigma) = M_{N-1}^*, \mathfrak{D}_N(\Sigma) = M_N^*$$

La proposition suivante donne une relation récursive pour la calculer.

Proposition III.4 (SV)

Quel que soit $S \in \mathcal{P}(\Sigma)$, et quel que soit $r = 2,3,\ldots,N$:

$$\overleftrightarrow{I}(\mathfrak{D}_r(S)) = \underset{\substack{\sigma \subseteq S \\ \emptyset \neq \sigma \neq S}}{\text{Min}} \{\overleftrightarrow{I}(\sigma : S \setminus \sigma) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S \setminus \sigma))\}$$

Démonstration

Cas $|S| < r$:

Dans ce cas le terme de gauche est par définition $+\infty$. Pour sa part, le terme de droite est aussi $+\infty$ puisque

$$\forall \sigma \subseteq S, \emptyset \neq \sigma \neq S : |S \setminus \sigma| < r-1$$

$$\Rightarrow \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S \setminus \sigma)) := +\infty$$

donc chacun des termes du Min est égale à $+\infty$.

Cas $|S| \geq r$:

Posons $\mathfrak{D}_r(S) = \{R_1^*, R_2^*, \ldots, R_r^*\} \in \mathbb{P}_r(S)$.
D'après la proposition II.12 ① (SV) nous avons :

$$\overleftrightarrow{I}(\mathfrak{D}_r(S)) = \overleftrightarrow{I}(R_1^*:R_2^*:\ldots:R_r^*)$$

$$= \overleftrightarrow{I}(R_1^*:R_2^* \cup R_3^* \cup \ldots \cup R_r^*) + \overleftrightarrow{I}(R_2^*:R_3^*:\ldots:R_r^*)$$

$$= \overleftrightarrow{I}(R_1^*:S \setminus R_1^*) + \overleftrightarrow{I}(R_2^*:R_3^*:\ldots:R_r^*)$$

Montrons maintenant que

$$\overleftrightarrow{I}(R_2^*:R_3^*:\ldots:R_r^*) = \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S \setminus R_1^*))$$

Supposons que ceci ne soit pas vrai, alors forcément

$$\overleftrightarrow{I}(R_2^*:R_3^*:\ldots:R_r^*) > \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S \setminus R_1^*))$$

(le contraire n'est pas possible puisque $\{R_2^*,R_3^*,\ldots,R_r^*\} \in \mathbb{P}_{r-1}(S \setminus R_1^*)$, et par définition $\mathfrak{D}_{r-1}(S \setminus R_1^*)$ est la partition optimale de $S \setminus R_1^*$, en r-1 classes).

Alors

$$\overleftrightarrow{I}(R_1^*:R_2^*:\ldots:R_r^*) = \overleftrightarrow{I}(R_1^*:S \setminus R_1^*) + \overleftrightarrow{I}(R_2^*:\ldots:R_r^*)$$

$$> \overleftrightarrow{I}(R_1^*:S \setminus R_1^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S \setminus R_1^*))$$

$$= \overleftrightarrow{I}(\{R_1^*\} \cup \mathfrak{D}_{r-1}(S \setminus R_1^*)),$$

mais cela constituerait une contradiction puisque $\{R_1^*\} \cup \mathfrak{D}_{r-1}(S \setminus R_1^*) \in \mathbb{P}_r(S)$, et on avait supposé $\{R_1^*,R_2^*,\ldots,R_r^*\}$ comme étant la partition optimale de S en r classes.

Ainsi donc

$$\overleftrightarrow{I}(R_2^*:R_3^*:\ldots:R_r^*) = \overleftrightarrow{I}(\mathfrak{D}_r(S \setminus R_1^*)),$$

c'est-à-dire

$$\overleftrightarrow{I}(\mathfrak{D}_r(S)) = \overleftrightarrow{I}(R_1^*:R_2^*:\ldots:R_r^*) = \overleftrightarrow{I}(R_1^*:S \setminus R_1^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S \setminus R_1^*)).$$

Montrons pour finir que

$$\overleftrightarrow{I}(R_1^*:S\setminus R_1^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus R_1^*)) = \underset{\substack{\sigma\subseteq S\\ \emptyset\neq\sigma\neq S}}{\text{Min}} \{\overleftrightarrow{I}(\sigma:S\setminus\sigma) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus\sigma))\}$$

Supposons que ceci ne soit pas vrai, alors forcément

$$\overleftrightarrow{I}(R_1^*:S\setminus R_1^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus R_1^*)) > \underset{\substack{\sigma\subseteq S\\ \emptyset\neq\sigma\neq S}}{\text{Min}} \{\overleftrightarrow{I}(\sigma:S\setminus\sigma) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus\sigma))\}$$

(Le contraire n'est pas possible puisque $R_1^* \subseteq S$ et $\emptyset \neq R_1^* \neq S$, donc le Min du terme de droite inclut le cas $\sigma = R_1^*$).
Supposons que ce Min est atteint pour $\sigma = \sigma^*$ ($\sigma^* \subseteq S$, $\emptyset \neq \sigma^* \neq S$).

Alors

$$\begin{aligned}\overleftrightarrow{I}(\mathfrak{D}_r(S)) &= \overleftrightarrow{I}(R_1^*:S\setminus R_1^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus R_1^*))\\ &> \overleftrightarrow{I}(\sigma^*:S\setminus\sigma^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus\sigma^*))\\ &= \overleftrightarrow{I}(\{\sigma^*\} \cup \mathfrak{D}_{r-1}(S\setminus\sigma^*)) .\end{aligned}$$

Mais cela constituerait une contradiction puisque

$\{\sigma^*\} \cup \mathfrak{D}_{r-1}(S\setminus\sigma^*) \in \mathbb{P}_r(S)$, et $\mathfrak{D}_r(S)$ est par définition la partition optimale de S en r classes.

Ainsi donc

$$\begin{aligned}\overleftrightarrow{I}(\mathfrak{D}_r(S)) &= \overleftrightarrow{I}(R_1^*:S\setminus R_1^*) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus R_1^*))\\ &= \underset{\substack{\sigma\subseteq S\\ \emptyset\neq\sigma\neq S}}{\text{Min}} \{\overleftrightarrow{I}(\sigma:S\setminus\sigma) + \overleftrightarrow{I}(\mathfrak{D}_{r-1}(S\setminus\sigma))\}\end{aligned}$$

(F.D.)

8.2 - ALGORITHME DE PROGRAMMATION DYNAMIQUE(PD)

En utilisant la proposition précédente, nous pouvons calculer les optimums locaux M_1^*, M_2^*,...,M_{N-1}^*,M_N^*.

En effet, pour r=2,3,...,N,

$$M_r^* = \mathfrak{D}_r(\Sigma) \Leftrightarrow \overleftrightarrow{I}(\mathfrak{D}_r(\Sigma)) = \underset{\substack{\sigma \subseteq \Sigma \\ \emptyset \neq \sigma \neq \Sigma}}{\text{Min}} \{\overleftrightarrow{I}(\sigma : \Sigma \setminus \sigma) + \underbrace{\overleftrightarrow{I}(\mathfrak{D}_{r-1}(\Sigma \setminus \sigma))}_{\text{"}\alpha\text{"}}\}$$

Cependant, on voit que pour calculer $\mathfrak{D}_r(\Sigma)$ il faut connaitre déjà $\mathfrak{D}_{r-1}(S)$, $\forall S \in \mathcal{P}(\Sigma)$, puisque l'argument du terme "α" parcourt tout l'ensemble $\mathcal{P}(\Sigma)$.

$\mathfrak{D}_1$ et $\mathfrak{D}_N$ étant trivialement connus, nous allons donc construire deux tableaux, D(.,.) et ID(.,.), qui ont la forme montrée dans la figure III.8. A l'intersection de la ligne $S(\forall S \in \mathcal{P}(\Sigma), |S| \geqslant 2)$, avec la colonne r ($2 \leqslant r \leqslant N-1$), l'information que ces tableaux contiennent est

D(S,r) = meilleure partition de S en r classes = $\mathfrak{D}_r(S)$;

ID (S,r) = Intervaluation Externe de la partition contenue dans D(S,r) = $\overleftrightarrow{I}(\mathfrak{D}_r(S))$.

En particulier, la colonne r=2 devra être calculée par une recherche exhaustive de la meilleure partition en deux de chacun des $S \in \mathcal{P}(\Sigma)$, $\emptyset \neq S \neq \Sigma$.

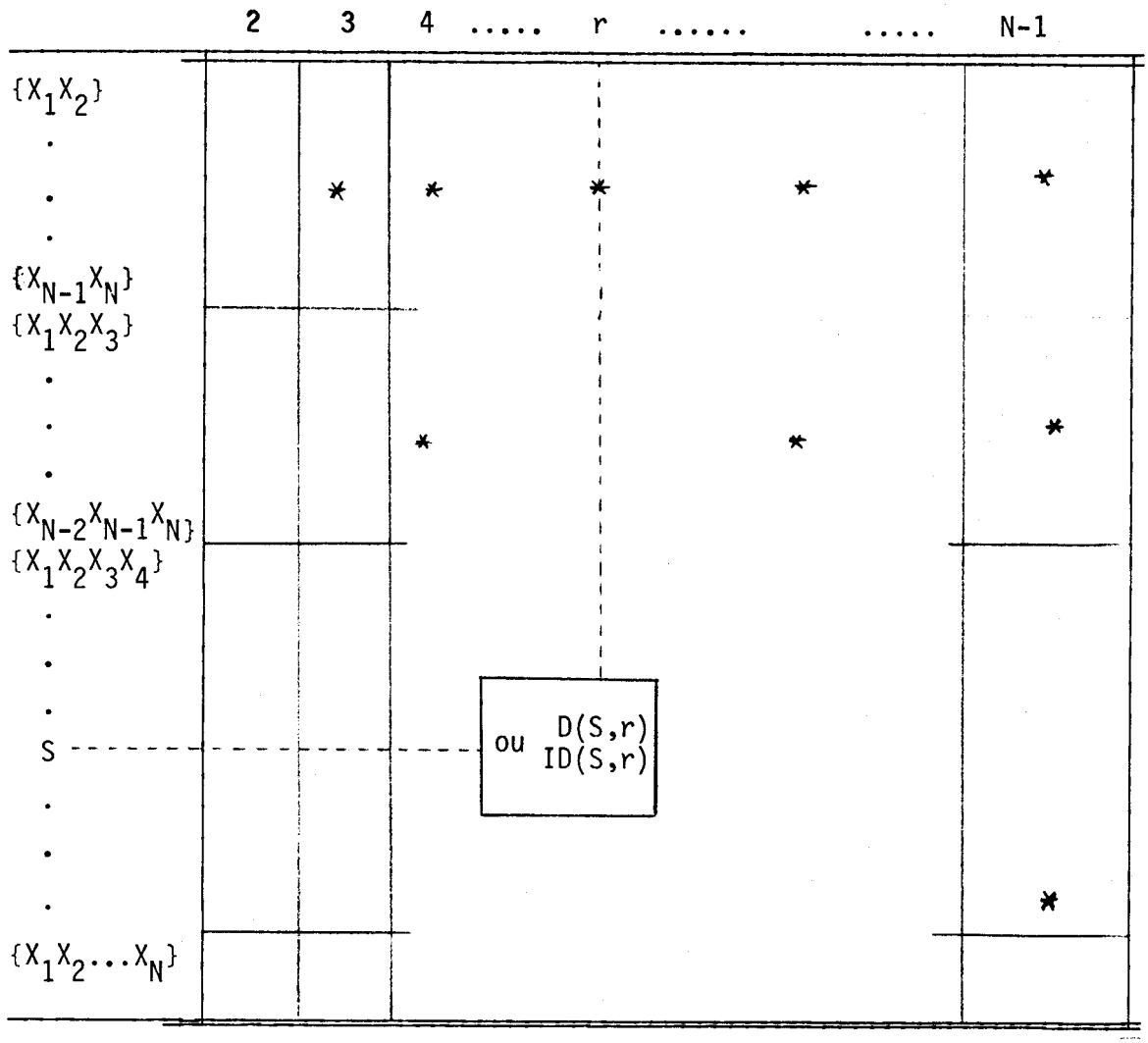


Figure III.8 Les tableaux D(.,.) et ID(.,.). Dans le tableau D(.,.), la partie supérieure à la "diagonale" reste vide. Dans le tableau ID(.,.), la partie supérieure à la "diagonale" contient $+\infty$.

L'algorithme de Programmation Dynamique est donné dans la page suivante.

```
.début;
ID (.,.) ← + ∞; (le tableau ID est initialisé à + ∞ partout)
Ecrire "{(X1,X2,...,XN)}",0;
Pour chaque S⊆Σ, |S|≥2 faire;
    Pour chaque σ⊆S, ∅≠σ≠S faire;                                   (*)
        Si I↔(σ:S\σ) < ID(S,2)
           alors;
              D(S,2) ← {σ,S\σ};
              ID(S,2) ← I↔(σ:S\σ);
        fsi;
    fpour;
fpour;
Ecrire D(Σ,2), ID(Σ,2);
r ← 3;
tant que r ≤ N-1 faire;
    Pour chaque S⊆Σ, |S| ≥ r faire;
        Pour chaque σ⊆S, ∅≠σ≠S faire;                               (**)
            Si I↔(σ:S\σ) + ID(S\σ,r-1) < ID(S,r)
               alors;
                  D(S,r) ← {σ,D(S\σ,r-1)};
                  ID(S,r) ← I↔(σ:S\σ) + ID(S\σ,r-1);
            fsi;
        fpour;
    fpour;
    Ecrire D(Σ,r),ID(Σ,r);
    r ← r+1;
ftq;
Ecrire "{(X1),(X2),...,(XN)}",I↔(X1:X2:...:XN);
fin;
```

REMARQUE : - Dans l'instruction marquée (*), il n'y a pas besoin d'examiner tous les $\sigma \subseteq S$, $\emptyset \neq \sigma \neq S$. Etant donné que $\overleftrightarrow{I}$ est symétrique, il suffit d'en examiner la moitié.

- Dans l'instruction marquée (**)il n'y a pas besoin non plus d'examiner tous les $\sigma \subseteq S$, $\emptyset \neq \sigma \neq S$. Il suffit d'examiner les $\sigma \subseteq S$, $1 \leqslant |\sigma| \leqslant |S| - r+1$. En effet, $S \setminus \sigma$ doit avoir au moins r-1 éléments pour qu'il possède des partitions en r-1 classes. Alors $|S \setminus \sigma| \geqslant r-1 \Rightarrow |S|-|\sigma| \geqslant r-1 \Rightarrow |\sigma| \leqslant |S|- r+1$

8.3 - OPTIMALITE DE L'ALGORITHME PD

Comme le montrent les paragraphes précédents, l'algorithme PD est optimal, dans le sens où il trouve tous les optimums locaux du treillis $\mathbb{P}(\Sigma)$. Mais pour cela il a fallu qu'il calcule aussi tous les optimums locaux des treillis $\mathbb{P}(S)$, $\forall S \subseteq \Sigma$, $\emptyset \neq S \neq \Sigma$. Ceci implique des calculs très longs, comme on va le montrer.

8.4 - ORDRE DE L'ALGORITHME PD

En toute rigueur, l'algorithme PD ne manipule pas tout à fait le même type d'objets que les algorithmes RE,AH,DR ou AS.

Pour ces derniers, leur démarche revient à l'examen d'un certain nombre de partitions de Σ "candidates" (pour chaque partition candidate il faut effectuer un petit nombre d'opération +,-, et quelques accès mémoire). Ce nombre de partitions candidates a été appelé "l'ordre de l'algorithme".

L'algorithme PD quant à lui possède deux boucles principales: la première détermine la meilleure partition en deux de chacun des $S \subseteq \Sigma$, $|S| \geqslant 2$, et pour cela il examine la moitié des $\sigma \subseteq S$, $\emptyset \neq \sigma \neq S$. La deuxième boucle détermine pour chaque $S \subseteq \Sigma$, $|S| \geqslant r (3 \leqslant r \leqslant N-1)$, sa meilleure partition en r classes, et pour cela il faut examiner tous les $\sigma \subseteq S$, $1 \leqslant |\sigma| \leqslant |S| - r+1$. Dans ce cas aussi, l'examen de chaque

σ requiert seulement un petit nombre d'opérations +,-, et d'accès mémoire. Nous allons donc calculer l'ordre de PD comme étant la somme du nombre des σ examinés dans la première boucle plus le nombre de ceux examinés dans la deuxième. Cela nous permettra d'avoir une base pour comparer le temps que cet algorithme nécessite avec celui de RE, AH, DR ou AS.

a) Ordre de la première boucle

Soit $S \subseteq \Sigma$, avec $|S| = r$. Le nombre de partitions $\{\sigma, S \setminus \sigma\}$ différentes est (c.f. § 5.3):

$$\frac{2^r - 2}{2} = 2^{r-1} - 1.$$

Par ailleurs, le nombre de sous-ensembles $S \subseteq \Sigma$, avec $|S| = r$ est

$$\binom{N}{r} := \frac{N!}{(N-r)!r!}$$

Ainsi donc, l'ordre de cette première boucle est donné par

$$\sum_{r=2}^{N} \binom{N}{r} (2^{r-1} - 1) = \frac{3^N - 2^{N+1} + 1}{2} \qquad [III.18]$$

b) Ordre de la deuxième boucle

Soit r donné ($3 \leq r \leq N-1$), et soit $S \subseteq \Sigma$, $|S| =: K \geq r$: le nombre de tels S est

$$\sum_{K=r}^{N} \binom{N}{r}.$$

Pour chacun de ces S, il faut examiner toutes les partitions de la forme $\{\sigma, S\setminus\sigma\}$, où $\sigma \subsetneq S$ et $1 \leqslant |\sigma| =: j \leqslant |S| - r + 1$: le nombre de telles partitions est

$$\sum_{j=1}^{K-r+1} \binom{K}{j}/2.$$

Ainsi donc, l'ordre de la deuxième boucle est

$$\sum_{r=3}^{N-1} \sum_{K-r}^{N} \binom{N}{K} \sum_{j=1}^{K-r+1} \frac{1}{2} \binom{K}{j}$$

c) Somme

L'ordre de l'algorithme PD est alors la somme des ordres des deux boucles, c'est-à-dire

$$\text{ordre (PD)} = \frac{3^N - 2^{N+1} + 1}{2} + \sum_{r=3}^{N-1} \sum_{K=r}^{N} \binom{N}{K} \sum_{j=1}^{K-r+1} \frac{1}{2} \binom{K}{j}$$

$$\overset{!!}{=} \sum_{r=2}^{N-1} \sum_{K=r}^{N} \binom{N}{j} \sum_{j+1}^{K-r+1} \frac{1}{2} \binom{K}{j}$$

$$= \frac{1}{2} \sum_{r=2}^{N-1} \sum_{K=r}^{N} \sum_{j=1}^{K-r+1} \frac{N!}{(N-K)\,!\,(K-j)\,!\,j\,!}$$

L'annexe 1 montre le calcul de ordre (PD) pour $|\Sigma| = N < 20$. On y voit que l'ordre de PD croit très vite, beaucoup plus que celui de AH, DR ou AS. Il est même curieux de remarquer que pour $2 \leqslant N \leqslant 9$ l'ordre de PD est supérieur à celui de l'algorithme de recherche exhaustive RE.

Comme pour les autres algorithmes, si l'on suppose que l'ordinateur peut exécuter mille fois par seconde les instructions d'une boucle, l'algorithme peut être envisagé jusqu'à $|\Sigma| = N \leqslant 14$. Dans ce cas l'ordre (PD) est 11'102.910, c'est-à-dire que le temps d'exécution serait 3 heures + 5 minutes + 2 secondes. Notons aussi que, à la différence des autres algorithmes, celui-ci requiert d'une quantité de mémoire non négligeable pour stocker ses deux tableaux D(.,.) et ID(.,.) (cette quantité est de l'ordre de $(N-1).(2^N-N-1)$).

9. CONCLUSIONS

Dans ce chapitre, nous avons étudié les partitions optimales, et nous avons vu la nécessité du concept d'optimums locaux. Nous avons aussi présenté cinq algorithmes.

- RE : Recherche Exhaustive,
- PD : Programmation Dynamique,

qui permettent de trouver tous les optimums locaux;

- AS : Approximations Successives,

qui trouvent les optimums locaux des niveaux 1 et N (trivialement), ainsi que ceux des niveaux 2 et N-1. Pour les autres niveaux les partitions proposées vérifient une condition nécessaire d'optimalité;enfin, les algorithmes.

- DR : Division Récursive,
- AH : Analyse Hiérarchique,

qui permettent de trouver des bonnes partitions, mais dont l'optimalité ne peut pas être caractérisée.

Pour avoir une idée du temps de calcul de chacun de ces algorithmes, nous avons étudié leur ordre. Avec ceux-ci nous pouvons établir des comparaisons, qui bien entendu n'ont qu'une valeur approximative. Ainsi, si un ordinateur peut réaliser 1000 fois/seconde les instructions d'une boucle principale (un petit nombre d'opérations +,-, et d'accés mémoire), et si l'on se fixe une limite de 2 1/2 heures de calcul, la taille maximale de Σ est donnée dans le tableau ci-dessous :

$\lvert\Sigma\rvert = N$ Maximal	RE	PD	AS	DR	AH
	12	13	20	23	377

Ces limites correspondent au cas où l'on parcourt les N niveaux du treillis $\mathbb{P}(\Sigma)$. On voit immédiatement que la méthode de Programmation Dynamique n'introduit pratiquement aucune amélioration par rapport à une Recherche Exhaustive. Quant aux algorithmes AS et DR, il faut rappeler que l'ordre que nous avons calculé constitue une borne largement supérieure, correspondant au pire des cas dans chacun des N niveaux.

Si par contre on se fixe d'avance le nombre r de sous-systèmes voulus, les algorithmes RE, PD et AS pourraient sans doute faire face à des systèmes plus grands.

Enfin, notons pour terminer que les algorithmes proposés sont valables quelle que soit la semi-valuation H utilisée, le calcul de $H(S)$, $\forall S \in \mathbb{P}(\Sigma)$, sera seulement plus ou moins lourd. Par ailleurs, la dualité de la représentation Σ^E et Σ^D, choisie au chapitre I, permet l'application de ces algorithmes aussi bien aux systèmes statiques qu'aux systèmes dynamiques. En particulier pour les systèmes dynamiques, l'hypothèse de déterminisme $H(\{X_1(t),X_2(t),\ldots,X_N(t)\}/\{X_1(t-1),X_2(t-1),\ldots,X_N(t-1)\})=0$, nécessaire aux approches proposés par [RICH 75,Chap.III] et [DUFO 79,Chap.II], n'a été utilisée nullepart.

CHAPITRE IV

DÉTERMINATION DES RELATIONS DANS UN SYSTÈME DE VARIABLES

I - INTRODUCTION

On peut en général diviser les variables d'un système en deux groupes : d'une part celles dont la mesure ne pose pas de problème particulier, et d'autre part, celles dont la mesure est difficile, coûteuse ou bruitée. Un autre cas fréquent est celui où l'on a besoin de prédire le comportement d'une variable (par exemple la variable "issue probable d'une opération") et où l'on dispose pour cela d'autres variables que l'on peut mesurer immédiatement (dossier médical d'entrée).

Des méthodes qui permettent d'extraire l'information qu'un groupe de variables peut fournir sur une autre variable seraient d'une très grande utilité pour ce type de situation. Le premier à proposer une méthodologie basée sur la théorie de l'Information et visant à déterminer les variables donnant le plus de renséignements sur une autre fût, à notre connaissance, R.E. RINK [1973]. Récemment R.C. CONANT [1980] proposa aussi une approche dans le même sens; cependant, peu ou pas de méthodes pratiques ont été proposées permettant la mise en oeuvre.

Dans ce chapitre nous reprenons les idées générales exposées par ces auteurs, et nous essayons d'affiner un peu plus les différents concepts, de présenter plusieurs méthodes concrêtes de mise en oeuvre, et d'explorer les applications à différents domaines.

On suppose qu'il a été fixé d'avance une division des variables de Σ en deux groupes : les *Variables à Expliquer*, et les *Variables Explicatives ou facilement observables*. On étudie d'abord le concept "explication d'une variable à partir d'autres", on définit celui de "Ensemble Optimal de Variables Explicatives", et ensuite on présente quatre algorithmes pour déterminer ou approximer ces ensembles optimaux. Pour chacun de ces algorithmes on étudie l'optimalité et le temps d'exécution requis.

Une fois connus les ensembles optimaux, on peut étudier leur application à l'Analyse Structurale. On introduit pour cela le concept "Graphe de Dépendance", sur lequel peuvent être appliquées les techniques courantes pour l'analyse de graphes.

Nous donnons ensuite un rapide aperçu de l'application des méthodes décrites à la Reconnaissance de Formes. Le chapitre se termine avec une courte discussion sur l'aide qui peut fournir la Théorie de l'Information à l'Identification explicite des relations ou équations liant les variables.

Enfin, rappelons que dans ce chapitre comme dans les précédents les interprétations et les motivations sont faites en termes de la Théorie de l'Information, mais les propositions et les algorithmes sont valables dans le contexte général indiqué par l'étiquette SV, SVM, SMP, ou SMPM qui les accompagne. S'il n'y a pas d'étiquette, le résultat ne concerne que le cas H = Entropie.

2 - EXPLIQUER UNE VARIABLE A PARTIR D'AUTRES

2.1 - INFORMATION COMMUNE A DEUX VARIABLES

Comme nous l'avons dit au chapitre II, la transinformation $\overleftrightarrow{I}(X:Y)$ mesure la quantité d'information commune aux variables X et Y. Cette redondance d'information est dûe en général à une transmission de données, ordres, paramètres, consignes, réponses, etc ... à travers un canal physique allant de X à Y et/ou vice-versa. Cependant, ce n'est pas le seul cas, et il est tout à fait possible (et même assez fréquent!) qu'il n'y ait pas de canal du tout entre X et Y, ni direct ni composé, et que pourtant $\overleftrightarrow{I}(X:Y)$ soit non négligeable [CONA 80]. Cela arrive lorsque d'autres variables influencent de la même façon les variables X et Y.

C'est par exemple le cas entre les variables
X(k): Nombre de voitures inmatriculées à Paris dans la période k, et
Y(k): Nombre de voitures inmatriculées a Londres dans la période k.
Les données statistiques des différentes années permettent de calculer que

$$\overleftrightarrow{I}(X:Y) \overset{\sim}{>} H(X) \overset{\sim}{>} H(Y) > 0 \qquad [IV.1]$$

alors qu'il est peu probable qu'il y ait des communications téléphoniques ou des échanges de correspondance suivis entre les deux bureaux d'immatriculation. L'explication est dûe au fait que d'autres variables comme l'inflation, le prix du pétrole, le taux d'échange du dollar, le taux d'intérêt aux Etats-Unis, le prix des matières premières, etc ..., influencent plus ou moins de la même façon le pouvoir d'achat des Anglais et des Français, donc leurs achats de véhicules.

Reprenant l'affirmation [IV.1], d'une façon informelle nous obtenons :

$$\begin{aligned} & H(X) \overset{\sim}{>} \overleftrightarrow{I}(X:Y) \\ \Rightarrow \quad & -H(X) \overset{\sim}{<} -\overleftrightarrow{I}(X:Y) := -H(X) - H(Y) + H(X,Y) \\ \Rightarrow \quad & 0 \overset{\sim}{<} H(X,Y) - H(Y) =: H(X/Y). \end{aligned} \qquad [IV.2]$$

Ainsi, si X ne nous était pas accessible, [IV.2] implique que connaissant Y et moyennant peut être quelques recodifications ou transformations, on serait en mesure de faire une bonne estimation de X (Nous discuterons l'identification explicite d'une telle fonction d'estimation au paragraphe 11).

Cet exemple illustre une autre caractéristique de la Transinformation, qui pour nos objectifs constitue un avantage supplémentaire : dans la recherche des variables servant à expliquer une variable X donnée, on ne sera pas limité aux variables liées physiquement à X.

2.2 - VARIABLES EXPLICATIVES ET VARIABLES A EXPLIQUER

Comme nous l'avons dit dans l'Introduction, les variables d'un système Σ peuvent en général être divisées en deux groupes : d'une part les *Variables à Expliquer*, c'est-à-dire celles dont la mesure est difficile, coûteuse, bruitée ou impossible à un instant donné; d'autre part les *Variables Explicatives ou facilement Observables*. Plusieurs cas peuvent se présenter, mais la formulation ci-dessous en inclut la plupart :

a) Systèmes Statiques

Soit $\Sigma^E = \{X_1^E, X_2^E, \ldots, X_N^E\}$ l'ensemble de toutes les variables du système (c.f. chap.I, § 4.2.1.). Moyennant une rénumérotation éventuelle, soit

$$\mathcal{E}_1^E := \{X_1^E, X_2^E, \ldots, X_e^E\} \qquad [IV.3]$$

l'ensemble des *Variables à Expliquer*. Pour chacune des $X_i^E \in \mathcal{E}^E$ on suppose qu'il a été donné un ensemble

$$\mathcal{O}_i^E \subseteq \Sigma^E \setminus \{X_i^E\} \qquad [IV.4]$$

de variables servant éventuellement à expliquer X_i^E. Cette condition [IV.4] empêche, bien entendu, que l'on se serve de X_i^E pour expliquer X_i^E.

b) Systèmes Dynamiques

On suppose ici qu'il faut expliquer certaines des variables du système à l'instant présent, et qu'on dispose pour cela de l'information fournie par les variables aux instants précédents. Voyons ceci formellement:

Soit $\Sigma^D = \{X_1^D, X_2^D, \ldots, X_N^D\}$ l'ensemble de toutes les variables du système (c.f. chap.I, § 4.2.2.) . Rappelons que Σ^D est sensé contenir les variables d'entrée du système (c.f. chap.I, §4.1). Chaque variable X_i^D est un vecteur

$$X_i^D := (X_i(t), X_i(t-1), X_i(t-2), \ldots, X_i(t-(p-1)))$$

qui regroupe les p dernières réalisations de la variable primaire X_i. Ecrivons Σ^D sous la forme suivante :

$$\Sigma^D = \{X_1(t), X_2(t), \ldots, X_N(t),$$
$$X_1(t-1), \ldots, X_1(t-(p-1)), X_2(t-1), \ldots, X_2(t-(p-1)), \ldots, X_N(t-1), \ldots, X_N(t-(p-1))\}$$

autrement dit, Σ^D contient toutes les variables à l'instant présent, décalées d'une période, décalées de 2 périodes, ..., décalées de p-1 périodes. Soit alors

$$\mathbb{E}^D := \{X_1(t), X_2(t), \ldots, X_e(t)\} \qquad [IV.5]$$

l'ensemble de *Variables à Expliquer*. Pour chaque $X_i \in \mathbb{E}^D$ on suppose qu'il a été donné un ensemble.

$$\sigma_i^D \subseteq \Sigma^D \setminus \{X_1(t), X_2(t), \ldots, X_N(t)\} \qquad [IV.6]$$
$$= \{X_1(t-1), \ldots, X_1(t-(p-1)), \ldots, X_N(t-1), \ldots, X_N(t-(p-1))\}$$

de variables servant éventuellement à expliquer X_i. Les variables d'entrée du système devraient en principe faire partie de ces σ_i^D.

Désormais nous utiliserons tout simplement $\mathcal{E}$ ou σ_i, avec $|\mathcal{E}|=e$, sans spécifier s'il s'agit du cas statique (E) ou dynamique (D). Les résultats donnés seront valables, bien entendu, quel que soit le cas.

2.3 - ENTROPIE (OU SEMI-VALUATION) DE X_i EXPLIQUEE PAR S

Si l'on prétend expliquer une variable $X_i \in \mathcal{E}$ par un sous-ensemble de variables $S \subseteq \sigma_i$, on a intérêt à choisir un S qui possède le maximum d'information commune avec X_i, c'est-à-dire, tel que $\overleftrightarrow{I}(X_i:S)$ soit aussi grand que possible.

Proposition IV.1 (SV)

Soit $X \in \mathcal{E}$. Quel que soit $S \subseteq \sigma_i$:

$$\overleftrightarrow{I}(X:S) + H(X/S) = H(X).$$

Interprétation

Quel que soit S, ce que S dit sur X plus ce qui reste à connaitre de X une fois connu S est égal à toute l'information que X peut fournir.

Démonstration :

$$\begin{aligned}&\overleftrightarrow{I}(X:S) + H(X/S)\\ &:=H(X) + H(S) - H(S,X) + H(S,X) - H(S)\\ &= H(X).\end{aligned}$$

(F.D.)

Ainsi, maximiser $\overleftrightarrow{I}(X:S)$ revient à minimiser $H(X/S)$. Nous allons donc travailler en termes de $H(X/S)$.

Définitions

Pour $X_i \in \mathcal{E}$ et $S \subseteq \sigma_i$:

* $\overleftrightarrow{I}(X_i:S) := H(X_i) + H(S) - H(S,X_i) = H(X_i) - H(X_i/S)$

est l'entropie (ou en général, la semi-valuation) de X_i expliquée par S.

* $H(X_i/S) := H(S,X_i) - H(S)$

est l'entropie (ou en général, la semi-valuation) de X_i non expliquée par S

* $\% H(X_i/S) := \dfrac{H(X_i/S)}{H(X_i)}$

est le pourcentage de l'entropie (ou en général, de la semi-valuation) de X_i non expliquée par S.

Proposition IV.2 (SMPM)

Quels que soient $X_i \in \mathcal{E}$ et $S \subseteq \sigma_i$ on a :

$$0 \leqslant H(X_i/\sigma_i) \leqslant H(X_i/S) \leqslant H(X_i)$$

ou ce qui est équivalent :

$$0 \leqslant \% H(X_i/\sigma_i) \leqslant \% H(X_i/S) \leqslant 1$$

Démonstration :

Etant donné que $\emptyset \subseteq S \subseteq \sigma_i$, l'inégalité énoncée est une conséquence immédiate de la proposition II.22 (SMPM).

(F.D.)

Proposition IV.3

Pour H = Entropie nous avons :

① $H(X/S) = 0$ (ou bien $\% H(X/S) = 0$)
ssi
il existe une fonction $f: M_S \longrightarrow M_X$ telle que $X = f(S)$;

② $H(X/S) = H(X)$ (ou bien $\% H(X/S) = 1$)
ssi

les variables X et S sont statistiquements indépendantes. (donc, il n'est pas question que S puisse servir à nous renseigner sur X)

Démonstration : Voir Annexe 4 (F.D.)

Comme le montrent les deux propositions précédentes, H(X/S) (respec.% H(X/S)) est compris entre 0 et H(X) (respec. 0 et 1), ces valeurs extrêmes étant atteintes dans des cas tout à fait opposés en ce qui concerne les possibilités pour que S nous renseigne sur X. Cela nous mène à considérer H(X/S) (respec. % H(X/S)) comme un indice (respect. un pourcentage) de *Non Modélisabilité de X à partir de S.*

Le plus grand ensemble de variables susceptibles d'expliquer ou de modéliser un $X_i \in \mathcal{E}$ étant par définition σ_i, nous pouvons donc considérer certaines limites :

** $H(X_i/\mathcal{O}_i) := H(\mathcal{O}_i, X_i) - H(\mathcal{O}_i)$

est la *Partie Non explicable* (ou non modélisable) de X_i. Autrement dit, c'est la partie de l'entropie (ou semi-valuation) de X_i qui provient soit d'un comportement autonome de X_i, soit de l'influence d'autres variables qui n'ont pas été inclues dans $\mathcal{O}_i$ (à ce titre, on pourrait aussi l'appeler "Bruit résiduel de X_i").

** $H(X_i) - H(X_i/\mathcal{O}_i) = \overleftrightarrow{I}(X_i : \mathcal{O}_i)$

est en conséquence la *Partie Explicable* (ou modélisable) de l'entropie (ou semi-valuation) de X_i.

** $\% H(X_i/\mathcal{O}_i) := \dfrac{H(X_i/\mathcal{O}_i)}{H(X_i)}$

est le *Pourcentage Non Explicable* (ou non modélisable) de l'entropie (ou semi-valuation) de X_i.

** $1 - \% H(X_i/\mathcal{O}_i) = \dfrac{\overleftrightarrow{I}(X_i : \mathcal{O}_i)}{H(X_i)}$

est en conséquence le *Pourcentage Explicable* (ou modélisable) de l'entropie (ou semi-valuation) de X_i.

Avec ces indices nous pouvons construire les vecteurs

$$\vec{H}(X./\mathcal{O}.) := \begin{pmatrix} H(X_1/\mathcal{O}_1) \\ H(X_2/\mathcal{O}_2) \\ \vdots \\ H(X_e/\mathcal{O}_e) \end{pmatrix} ; \% \vec{H}(X./\mathcal{O}.) := \begin{pmatrix} \% H(X_1/\mathcal{O}_1) \\ \% H(X_2/\mathcal{O}_2) \\ \vdots \\ \% H(X_e/\mathcal{O}_e) \end{pmatrix} \qquad [IV.7]$$

Considérons une norme du type Minkouski du vecteur $\vec{H}(X./\mathcal{G}.)$

$$|| \vec{H}(X./\mathcal{G}.)||_{\alpha} := \left[\sum_{i=1}^{e} H(X_i/\mathcal{G}_i)^{\alpha}\right]^{1/\alpha} , \qquad \alpha \geq 1 \qquad [IV.8]$$

on obtient alors une indication du total non explicable des variables de $\mathcal{E}$. Il nous semble que le plus naturel, tout au moins pour H = Entropie, est de prendre $\alpha = 1$. On définit donc le *Total Non Explicable* TNE ainsi :

$$TNE(X./\mathcal{G}.) := || \vec{H}(X./\mathcal{G}.) ||_1 := \sum_{i=1}^{e} H(X_i/\mathcal{G}_i) \qquad [IV.9]$$

on montre immédiatement, d'après la proposition II.17 (SV), que

$$(SV) : TNE\ (X./\mathcal{G}.) := \sum_{i=1}^{e} H(X_i/\mathcal{G}_i) \leq \sum_{i=1}^{e} H(X_i), \qquad [IV.10]$$

ce qui nous permet de définir le *Pourcentage Total Non Explicable* :

$$\% \text{ TNE } (X./\mathcal{G}.) := \frac{TNE(X./\mathcal{G}.)}{\sum_{i=1}^{e} H(X_i)} \qquad [IV.11]$$

2.4 - LE TOTAL NON EXPLICABLE : UN INDICE IMPORTANT

2.4.1. - Pour qualifier le problème d'estimation

Cet indice % TNE $(X./\mathcal{G}.)$ peut nous donner une idée de la faisabilité du problème "Estimer chaque $X_i \in \mathcal{E}$ en utilisant les variables de $\mathcal{G}_i$", c'est-à-dire, s'il sera possible d'arriver à un résultat de la forme

$$X_i \overset{\sim}{\sim} f_i(\theta_i) \qquad i = 1,2,\ldots,e$$

qui soit acceptable. En effet, si % TNE ($X./\theta.$) à une valeur proche de zéro, cela veut dire qu'il y a suffisamment de redondance dans le système, ou bien, que les relations de causalité entre variables explicatives et variables à expliquer sont fortes. Par contre, une valeur proche de 1 de % TNE ($X./\theta.$) indique qu'il y a trop de variables à expliquer, et/ou trop peu de variables explicatives, ou tout simplement, qu'il y a peu de chances d'établir un modèle mathématique servant à estimer les variables de ξ vu le manque d'informations disponibles.

2.4.2. - Pour qualifier le choix de Σ

L'application des techniques de Contrôle, Optimisation, Commande Optimale, ..., à un système requiert que l'on dispose d'un modèle mathématique du système. Or la possibilité de construire un tel modèle dépend de la mesure dans laquelle le système "s'explique en lui même", c'est-à-dire, dans quelle mesure n'est-on pas obligé d'aller chercher des explications dans des variables ou objets externes à Σ (rappelons que Σ est sensé contenir les variables d'entrée du système (c.f. chap.I, § 4.1)). Examinons séparément les cas statique et dynamique.

a) Cas statique

Quelle que soit la variable $X_i^E \in \Sigma^E$, la quantité % $H(X_i/\theta_i)$ indique la possibilité d'obtenir une relation de la forme

$$X_i = f_i(\theta_i^E) \text{ , avec } \theta_i^E \subseteq \Sigma^E \setminus \{X_i^E\}. \qquad [IV.12]$$

Nous pouvons dire alors que pour tout sous-ensemble $\xi^E \subseteq \Sigma^E$, avec $|\xi^E| = p$, la valeur

$$\frac{\sum_{X_i^E \in \xi^E} H(X_i^E/\theta_i^E)}{\sum_{X_i^E \in \xi^E} H(X_i^E)} =: \% \text{ TNE } (\xi^E|\theta^E) \qquad [IV.13]$$

indique la possibilité d'établir p équations entre les N variables de Σ, autrement dit, de construire un modèle de Σ à N-p degrés de liberté (au minimum), les variables de $\mathfrak{E}^E$ étant les variables dépendantes.

Alors, pour un système statique à N-p degrés de liberté, la valeur

$$\text{Min } \{ \% \text{ TNE } (\mathfrak{E}^E | \mathfrak{O}^E) : \mathfrak{E}^E \subseteq \Sigma^E \text{ et } |\mathfrak{E}^E| = p\} \qquad [IV.13]$$

qualifie la pertinence du choix des N variables de Σ^E. En effet, cette valeur représente le total non expliqué par le meilleur modèle possible de Σ^E ayant au moins N-p dégrés de liberté.

En particulier, le modèle avec le moins de degrés de liberté possible est celui où l'on prend

$$\mathfrak{E}^E = \{ X_1^E, X_2^E, \ldots, X_N^E\}$$

$$\mathfrak{O}_i^E = \Sigma^E \setminus \{X_i^E\} \qquad [IV.14]$$

car dans ce cas on cherche à établir N équations entre les N variables.

Notons enfin que avec $\mathfrak{O}_i^E$ défini par [IV.14] nous pouvons obtenir un résultat un peu plus fin que l'affirmation [IV.10] (SV):

Proposition IV.4 (SMPM)

Soit $\mathfrak{E}^E := \{X_1^E, X_2^E, \ldots, X_e^E\}$, et soit $\mathfrak{O}_i^E := \Sigma^E \setminus \{X_i^E\}$. Alors :

$$\text{TNE } (X_\cdot^E / \mathfrak{O}_\cdot^E) := \sum_{i=1}^{e} H(X_i^E / \mathfrak{O}_i^E) \leq H(X_1^E, X_2^E, \ldots, X_e^E) =: H(\mathfrak{E}^E).$$

Démonstration :
==========

Voir Annexe 5. (F.D.)

b) Cas Dynamique

On redéfinit

$$E^D := \{X_1(t), X_2(t), \ldots, X_e(t)\} \qquad [IV.15]$$

avec $\{X_{e+1}(t), \ldots X_N(t)\}$ étant l'ensemble de variables d'entrée. Soit

$$\mathcal{O}_i^D := \Sigma^D \setminus \{X_1(t), X_2(t), \ldots, X_N(t)\} \qquad [IV.16]$$

L'indice % TNE ($X./\mathcal{O}.$) nous indique alors s'il est possible d'obtenir un résultat de la forme

$$X_i(t) = f_i(X_1(t-1), \ldots, X_N(t-1), \ldots, X_1(t-(p-1)), \ldots, X_N(t-(p-1))) \qquad [IV.17]$$

$$1 \leqslant i \leqslant e,$$

autrement dit, d'obtenir un modèle mathématique qui exprime chaque variable à l'instant t en fonction des variables et des entrées aux instants précédents.

Dans les deux cas, une valeur élevée de % TNE ($X./\mathcal{O}.$) nous indiquerait qu'on a laissé quelques variables importantes en dehors de Σ (figure IV.1), ou/et que l'ordre p choisi (c.f. chap.I, § 4.2.2.) est insuffisant (cas dynamique), ou tout simplement, qu'il n'est pas possible de construire un modèle déterministe;

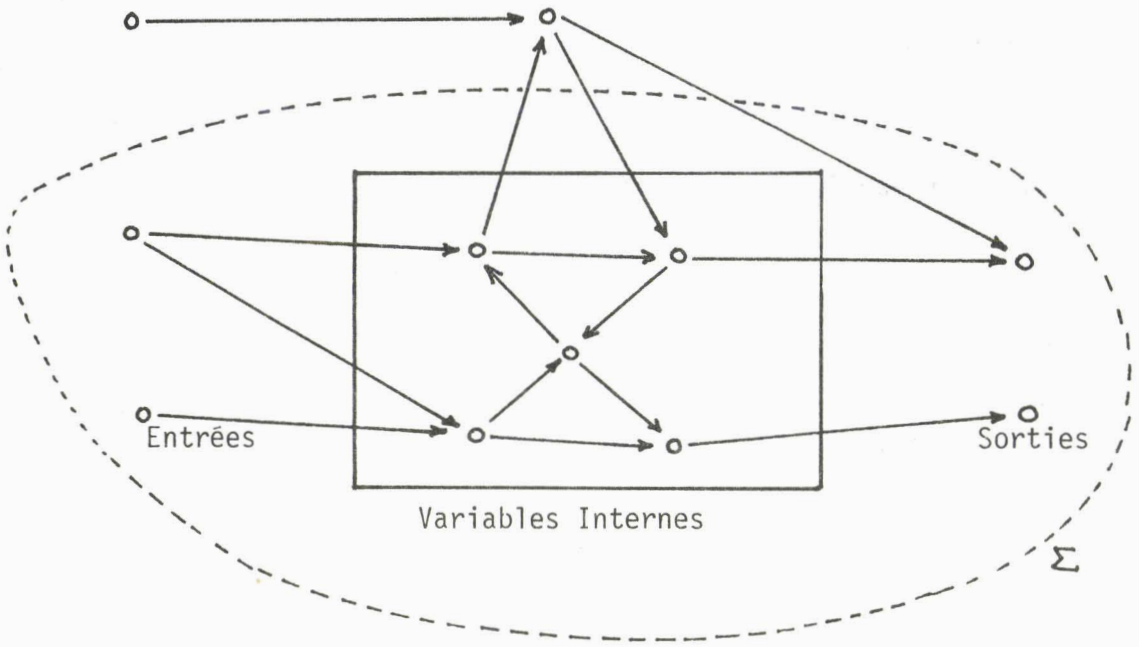


Figure IV.1 Σ incomplet

Inversement, une petite valeur de % TNE (X./θ.) indiquerait que Σ contient toutes les variables importantes (figure IV.2) et qu'il est possible de construire un modèle déterministe.

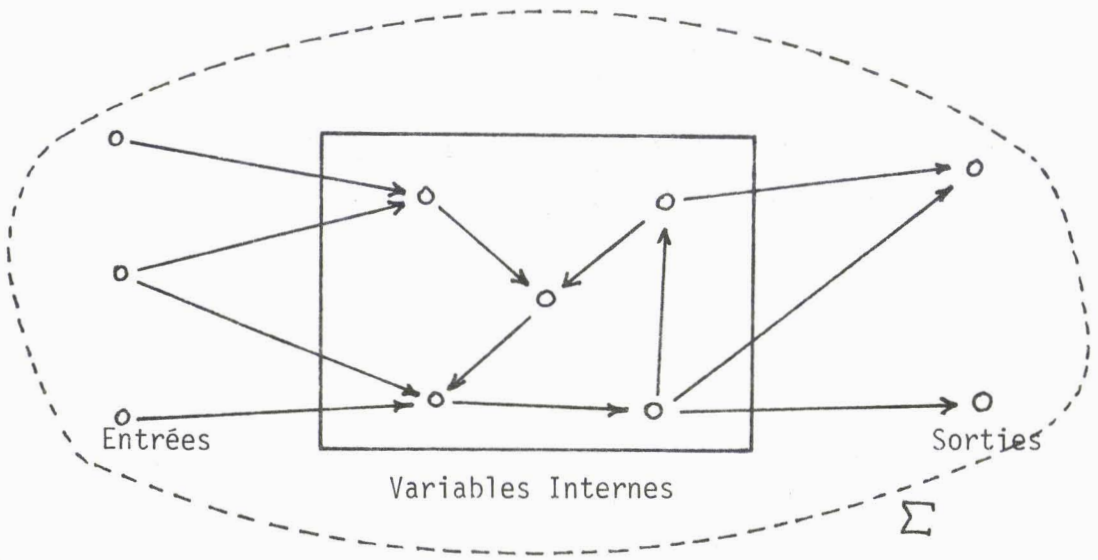


Figure IV.2 Σ Complet

Il faut néanmoins remarquer qu'on n'a fait aucune hypothèse (linéarité, continuité, ...) sur les f_i mentionnés dans [IV.12] et [IV.17]. Il s'agit seulement de trouver des relations univoques f_i entre les variables $\mathcal{O}_i$ et X_i. Au paragraphe 11 nous parlerons du cas où l'on contraint les f_i à appartenir à une classe de fonctions donnée.

2.5 - PARTIE EXPLIQUEE DE LA PARTIE EXPLICABLE

Nous disposons déjà des indices $H(X_i/S)$ et $\%\ H(X_i/S)$ pour qualifier le choix de S vis à vis de X_i. Cependant, une affirmation telle que $H(X_i/S) = 7$ ou $\%\ H(X_i/S) = 13\%$ est encore ambigüe étant donné les différences d'échelle, et même pour l'indice normalisé $\%\ H(X_i/S)$, du fait que rien n'est dit sur, la possibilité d'améliorer encore le choix de S. Pour remédier à ce problème nous proposons l'indice *Pourcentage Encore Non expliqué par S de la Partie Explicable de* X_i ($\%$ ENEPE (X_i/S)), défini par :

$\forall X_i \in \mathcal{X}$; $\forall S \subseteq \mathcal{O}_i$:

$$\%\ \text{ENEPE}\ (X_i/S) := \begin{cases} \dfrac{H(X_i/S) - H(X_i/\mathcal{O}_i)}{\overleftrightarrow{I}(X_i : \mathcal{O}_i)} & \text{si } \overleftrightarrow{I}(X_i:\mathcal{O}_i) > 0 \\ 0 & \text{si } \overleftrightarrow{I}(X_i:\mathcal{O}_i) = 0 \end{cases} \qquad [IV.18]$$

Interprêtons intuitivement cet indice :

$$\frac{H(X_i/S) - H(X_i/\mathcal{O}_i)}{\overleftrightarrow{I}(X_i:\mathcal{O}_i)} =$$

$$= \frac{\left[\begin{array}{l}\text{Partie de } X_i \text{ qui reste encore} \\ \text{à expliquer sachant S}\end{array}\right] - \left[\text{Partie non explicable de } X_i\right]}{\left[\text{Partie explicable de } X_i\right]}$$

$$= \frac{\left[\text{Partie de } X_i \text{ qu'il serait encore possible d'expliquer}\right]}{\left[\text{Partie explicable de } X_i\right]}$$

$$= \left[\text{Pourcentage Encore non Expliqué par S de la Partie Explicable de } X_i\right]$$

Par ailleurs, si $\overleftrightarrow{I}(X_i : \mathcal{C}_i) = 0$, la partie explicable de X_i est nulle, dont trivialement, il ne resterait rien à expliquer de cette partie explicable.

Proposition IV.5 (SMPM)

Soit $\overleftrightarrow{I}(X_i : \mathcal{C}_i) > 0$. Alors $\forall X_i \in \mathcal{E}$ et $\forall S \subseteq \mathcal{C}_i$:

$$0 \leq \%\ \text{ENEPE}\ (X_i/S) \leq 1,$$

avec

$$\%\ \text{ENEPE}\ (X_i/S) = 0 \Leftrightarrow H(X_i/S) = H(X_i/\mathcal{C}_i),$$

et

$$\%\ \text{ENEPE}\ (X_i/S) = 1 \Leftrightarrow H(X_i/S) = H(X_i).$$

Démonstration :

Etant donné que $S \subseteq \mathcal{C}_i$, d'après la proposition IV.2 (SMPM) nous avons :

$$H(X_i/S) \geqslant H(X_i/\mathcal{O}_i)$$

$$\Rightarrow \quad H(X_i/S) - H(X_i/\mathcal{O}_i) \geqslant 0$$

$$\Rightarrow \quad \frac{H(X_i/S) - H(X_i/\mathcal{O}_i)}{\overrightarrow{I}(X_i:\mathcal{O}_i)} \geqslant 0.$$

D'autre part, d'après la proposition II.17 (SV), nous avons

$$H(X_i/S) \leqslant H(X_i)$$

$$\Rightarrow \quad H(X_i/S) - H(X_i/\mathcal{O}_i) \leqslant H(X_i) - H(X_i/\mathcal{O}_i) = \overleftrightarrow{I}(X_i:\mathcal{O}_i)$$

$$\Rightarrow \quad \frac{H(X_i/S) - H(X_i/\mathcal{O}_i)}{\overrightarrow{I}(X_i:\mathcal{O}_i)} \leqslant 1.$$

Les deux autres affirmations sont immédiates (F.D.)

Remarquons pour terminer que les trois indices $H(X_i/S)$, % $H(X_i/S)$ et % ENEPE (X_i/S), permettant de qualifier le choix d'un S vis à vis de X_i, ne sont qu'un essai de formalisation des concepts de "partie modélisable", "pourcentage explicable", ... Mais en réalité, comme le montre la proposition ci-dessous, ils conduisent tous les trois au même résultat.

Proposition IV.6 (SMPM)

Soit $X_i \in \mathcal{E}$, et soient S et S' deux sous-ensembles de $\mathcal{O}_i$. Les trois affirmations suivantes sont équivalentes:

$$\Leftrightarrow \quad \text{i)} \quad H(X_i/S) \leqslant H(X_i/S')$$

$$\Leftrightarrow \quad \text{ii)} \quad \%\ H(X_i/S) \leqslant \%\ H(X_i/S')$$

$$\Leftrightarrow \quad \text{iii)} \quad \%\ \text{ENEPE}\ (X_i/S) \leqslant \%\ \text{ENEPE}\ (X_i/S').$$

Démonstration :
==========

Si $\overleftrightarrow{I}(X_i:\Theta_i) = 0$, alors $H(X_i) = H(X_i/S) = H(X_i/S') = H(X_i/\Theta_i)$, et les trois affirmations sont trivialement vérifiées.

Soit maintenant $\overleftrightarrow{I}(X_i:O_i) > 0$. Il est facile de voir que ii $\Leftrightarrow$ i $\Leftrightarrow$ iii :

(ii) $$\frac{H(X_i/S)}{H(X_i)} =: \%\, H(X_i/S) \leqslant \%\, H(X_i/S') := \frac{H(X_i/S')}{H(X_i)}$$

$\Leftrightarrow$

(i) $$H(X_i/S) \leqslant H(X_i/S')$$

$\Leftrightarrow$

$$H(X_i/S) - H(X_i/\Theta_i) \leqslant H(X_i/S') - H(X_i/\Theta_i)$$

$\Leftrightarrow$

(iii) $$\%\ \mathrm{ENEPE}(X_i/S) := \frac{H(X_i/S)-H(X_i/\Theta_i)}{\overleftrightarrow{I}(X_i : \Theta_i)} \leqslant \frac{H(X_i/S')-H(X_i/\Theta_i)}{\overleftrightarrow{I}(X_i : \Theta_i)} =: \%\mathrm{ENEPE}(X_i/S')$$

(F.D.)

3 - SOUS-ENSEMBLES EXPLICATIFS OPTIMAUX

3.1 - DEFINITION DES OPTIMUMS LOCAUX

Notre objectif est de trouver pour chaque $X_i \in \mathcal{E}$, le sous-ensemble $\xi_i \subseteq \mathcal{O}_i$ tel que

$$H(X_i/\xi_i) \leq H(X_i/S), \quad \forall S \subseteq \mathcal{O}_i \qquad [IV.19]$$

Cependant, le problème posé de cette façon conduit, comme le montre la proposition IV.2 (SMPM), à la solution triviale

$$\xi_i = \mathcal{O}_i, \text{ pour } i = 1,2,\ldots,e.$$

Or, cette solution triviale présente deux inconvénients : d'abord, elle ne nous fournit aucun renseignement sur la structure du système, puisque elle revient à dire "une variable est expliquée par (donc "dépendante de", "influencée par", "branchée à", ...) toutes les variables susceptibles de l'expliquer". Mais surtout, le problème qui se pose ensuite celui d'identifier la fonction f_i telle que $X_i \approx f_i(\xi_i)$, augmente inutilement de dimension.

Ainsi, comme pour le problème des partitions optimales, nous allons plutôt chercher des optimuls locaux. Le problème est alors énoncé sous la forme suivante :

> $\forall X_i \in \mathcal{E}$, et pour $K = 1,2,\ldots, |\mathcal{O}_i|$, trouver
>
> $\xi_i^K \subseteq \mathcal{O}_i$, avec $|\xi_i^K| = K$, tel que
>
> $\forall S \subseteq \mathcal{O}_i$, $|S| = K$: $H(X_i/\xi_i^K) \leq H(X_i/S)$. [IV.20]

On voit immédiatement que pour tout $X_i \in \mathcal{E}$:

$$* \quad \xi_i^0 = \emptyset \qquad\qquad * \quad \xi_i^{|\mathcal{O}_i|} = \mathcal{O}_i .$$

Appelons $\xi_\cdot^K := (\xi_1^K, \xi_2^K, \ldots, \xi_e^K)$.

Proposition IV.7 (SMPM)

Quel que soit $X_i \in \mathcal{E}$, et quel que soit K, $1 \leqslant K \leqslant |\mathcal{O}_i|$:

$$H(X_i/\xi_i^K) \leqslant H(X_i/\xi_i^{K-1})$$

Démonstration :

Soit $X_\alpha \in \mathcal{O}_i \setminus \xi_i^{K-1}$, et soit $A := \xi^{K-1} \cup \{X_\alpha\}$. Alors $\xi_i^{K-1} \subseteq A$, et d'après la Proposition II.22 (SMPM) on obtient

$$H(X_i/\xi_i^{K-1}) \geqslant H(X_i/A)$$

D'autre part, $|A| = K$, donc par définition de ξ_i^K :

$$H(X_i/\xi_i^K) \leqslant H(X_i/A)$$

De ces deux inégalités on déduit

$$H(X_i/\xi_i^K) \leqslant H(X_i/\xi_i^{K-1}) \qquad \text{(F.D.)}$$

Comme conséquence directe des propositions IV.6 et IV.7 on a les quatre affirmations suivantes (valables pour H = SMPM) :

A1)

$$1=\% H(X_i/\xi_i^0) \geq \% H(X_i/\xi_i^1) \geq \% H(X_i/\xi_i^2) \geq \ldots \geq \% H(X_i/\xi_i^{|\sigma_i|}) = H(X_i/\sigma_i) \quad [IV.21]$$

A2)

$$1=\% ENEPE(X_i/\xi_i^0) \geq \% ENEPE(X_i/\xi_i^1) \geq \ldots \geq \% ENEPE(X_i/\xi_i^{|\sigma_i|}) = 0. \quad [IV.22]$$

Si $\forall X_i \in E$ on a $|\sigma_i| = c$, nous pouvons dire par ailleurs que

A3)

$$\sum_{i=1}^{e} H(X_i)=TNE(X./\xi_.^0) \geq TNE(X./\xi_.^1) \geq TNE(X./\xi_.^2) \geq \ldots \geq TNE(X./\xi_.^c) = TNE(X./\sigma.) \quad [IV.23]$$

A4)

$$1= TNE(X./\xi_.^0) \geq \% TNE(X./\xi_.^1) \geq TNE(X./\xi_.^2) \geq \ldots \geq \% TNE(X./\xi_.^c) =\% TNE(X./\sigma.). \quad [IV.24]$$

Il faut toutefois remarquer que les affirmations précédentes n'impliquent pas que $\xi_i^0, \xi_i^1, \xi_i^2, \ldots \xi_i^{|\sigma_i|}$ constitue une suite croissante, c'est-à-dire, que ξ_i^K n'est pas forcément contenu dans ξ_i^{K+1} .

3.2 - AVANTAGES DE L'APPROCHE PAR OPTIMUMS LOCAUX

Cette approche par Optimums Locaux présente deux avantages :

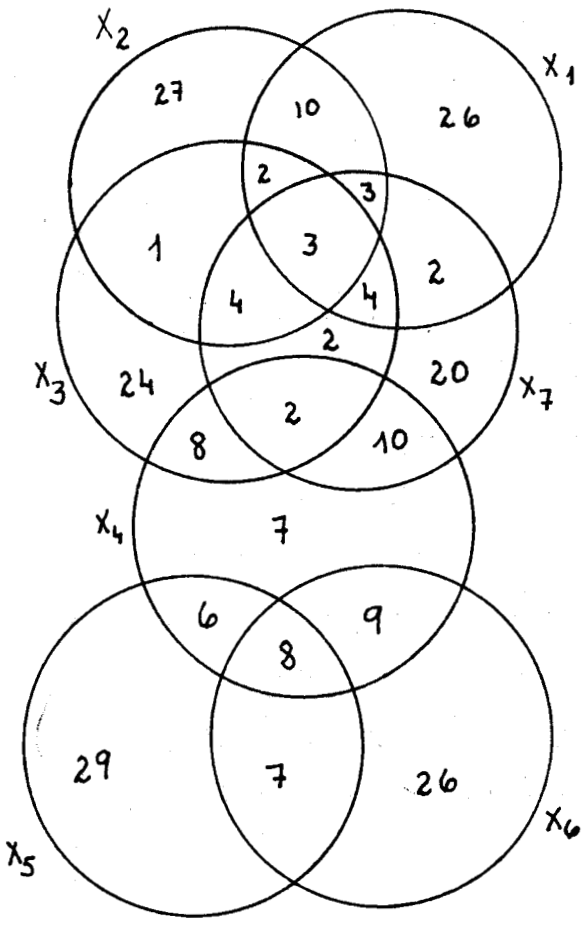
i) Comme le montrent la proposition IV.7 (SMPM) et les affirmations [IV.21] à [IV.24], les $\xi_.^K := (\xi_1^K, \ldots, \xi_e^K)$, K = 0,1,2,... constituent une hiérarchie de solutions de plus en plus exactes. Le praticien peut alors faire un compromis entre l'exactitude de l'estimation et la dimension du modèle qui serait nécessaire.

ii) La connaissance des $\xi^K_\cdot$, K = 0,1,2,..., nous fournit des renseignements précieux sur la structure du système. En effet pour chaque variable $X_i \in \Sigma$ on obtient une hiérarchie ξ^1_i, ξ^2_i,..., $\xi^{|\mathcal{O}_i|}_i$, des sous-ensembles qui expliquent (ou influencent) le mieux X_i. Mais surtout, la méthode utilisée est *multidimensionnelle*, dans le sens où elle tient compte de l'information que toutes les variables de ξ^K_i *prises dans leur ensemble* fournissent sur X_i, ce qui peut être très différent d'une addition de l'information que chacune des variables de ξ^K_i fournit séparément sur X_i. Dans le Chapitre II, paragraphe 9.3, nous avons présenté un système $\Sigma = \{X_1, X_2, X_3, X_4\}$ ou l'information apportée par X_1 ou par X_2 ou par X_3, sur X_4 est nulle, et ce n'est qu'en prennant X_1 et X_2 et X_3 ensemble que l'on trouve qu'elles expliquent complètement le comportement de X_4.

4 - UN EXEMPLE D'APPLICATION DES INDICES DÉFINIS

Nous reprenons le système $\Sigma = \{X_1, X_2, \ldots, X_7\}$, qui a été présenté au chapitre III, §7. Bien sur, il ne s'agit que d'un exemple purement "académique", mais qui permet d'illustrer le comportement des indices présentés auparavant.

Soit donc $\Sigma = \{X_1, X_2, \ldots, X_7\}$, dont la représentation graphique est montrée dans la figure IV.3.



$H(X_i)=50 \quad i=1,2,\ldots,7$

$H(\Sigma)=240$

$\sum_{i=1}^{7} H(X_i)=350$

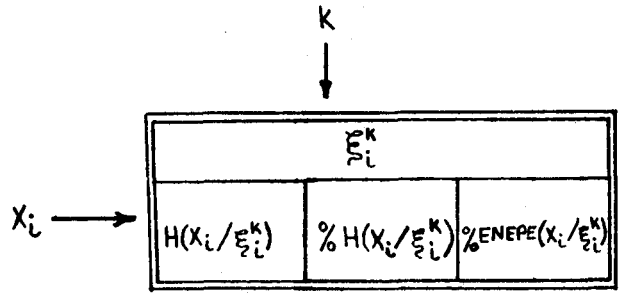
$\overleftrightarrow{I}(X_1:X_2:\ldots:X_7)=110$

$TNE(X./\mathcal{O}.)=159$

$\%TNE(X./\mathcal{O}.)=.662 \times 100\%$

Figure IV.3 Représentation graphique de $\Sigma=\{X_1,\ldots,X_7\}$

Soit $E := \{X_1, X_2, \ldots, X_7\}$, et $\sigma_i := \Sigma \setminus \{X_i\}$, $1 \leq i \leq 7$.
Le tableau ci-dessous montre les différents résultats



X_i ↓ K →

X_i	1			2			3			4			5			6			$H(X_i/\sigma_i)$	$\%H(X_i/\sigma_i)$
X_1	X_2			X_2, X_7			"			"			"			"			26	.52
	32	.64	.25	26	.52	0	"	"	"	"	"	"	"	"	"	"	"	"		
X_2	X_1			X_1, X_3			"			"			"			"			27	.54
	32	.64	.21	27	.54	0	"	"	"	"	"	"	"	"	"	"	"	"		
X_3	X_7			X_4, X_7			X_2, X_4, X_7			"			"			"			24	.48
	35	.70	.42	27	.54	.11	24	.48	0	"	"	"	"	"	"	"	"	"		
X_4	X_6			X_6, X_7			X_3, X_6, X_7			X_3, X_5, X_6, X_7			"			"			7	.14
	33	.66	.60	21	.42	.32	13	.26	.13	7	.14	0	"	"	"	"	"	"		
X_5	X_6			X_4, X_6			"			"			"			"			29	.58
	35	.70	.28	29	.58	0	"	"	"	"	"	"	"	"	"	"	"	"		
X_6	X_4			X_4, X_5			"			"			"			"			26	.52
	33	.66	.29	26	.52	0	"	"	"	"	"	"	"	"	"	"	"	"		
X_7	X_3			X_3, X_4			X_1, X_3, X_4			"			"			"			20	.40
	35	.70	.50	25	.50	.16	20	.40	0	"	"	"	"	"	"	"	"	"		
$TNE(X./\xi_.^K)$	235			181			165			159			159			159				
$\%TNE(X./\xi_.^K)$	.671			.517			.471			.454			.454			.454				

Les figures suivantes montrent les graphiques des indices $H(X_i/\xi_i^K)$, $\%\, H(X_i/\xi_i^K)$, $\%$ ENEPE (X_i/ξ_i^K), $K=1,2,\ldots,6$, ainsi que des indices $TNE(X./\xi_.^K)$ et $\%\, TNE(X./\xi_.^K)$. On y voit en particulier qu'à partir d'un certain K (le premier pour lequel $\%\, ENEPE(X_i/\xi_i^K) = 0$), les résultats ne peuvent plus être améliorés. Ce n'est pas la peine d'aller plus loin.

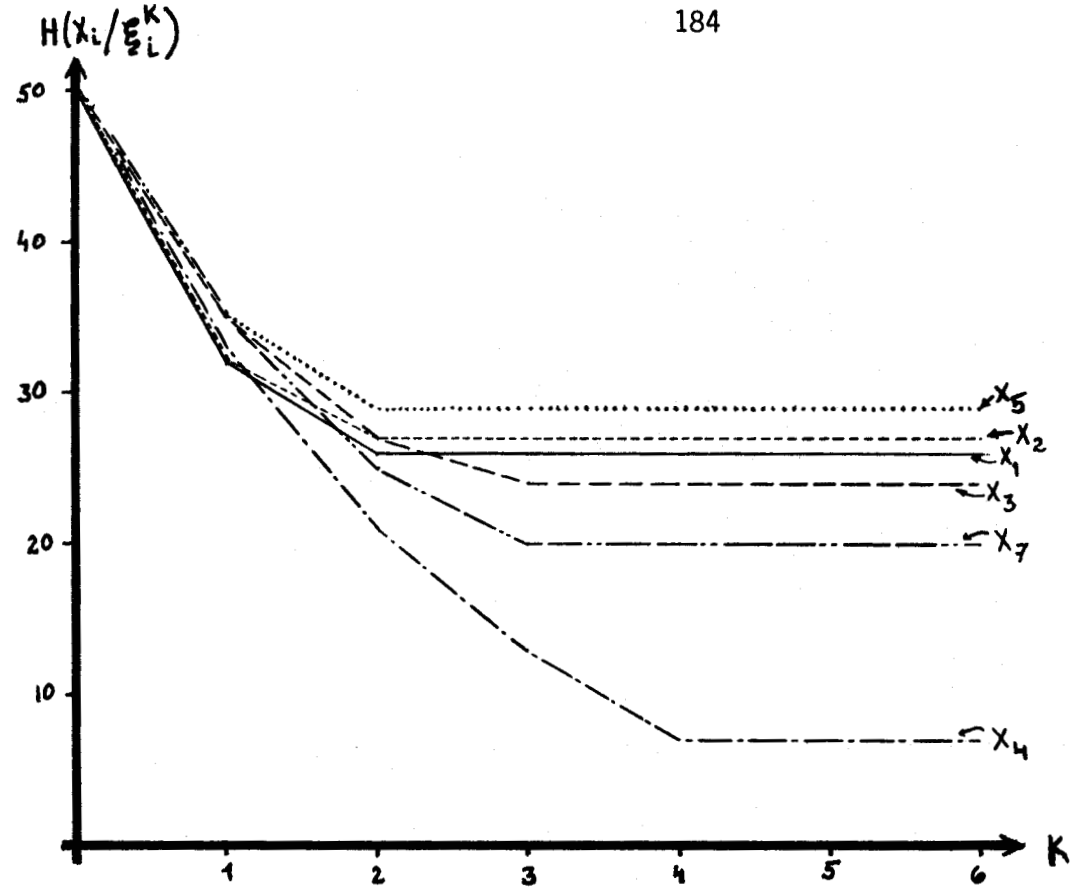


Figure IV.4 Partie Non Expliquée de X_i sachant ξ_i^k

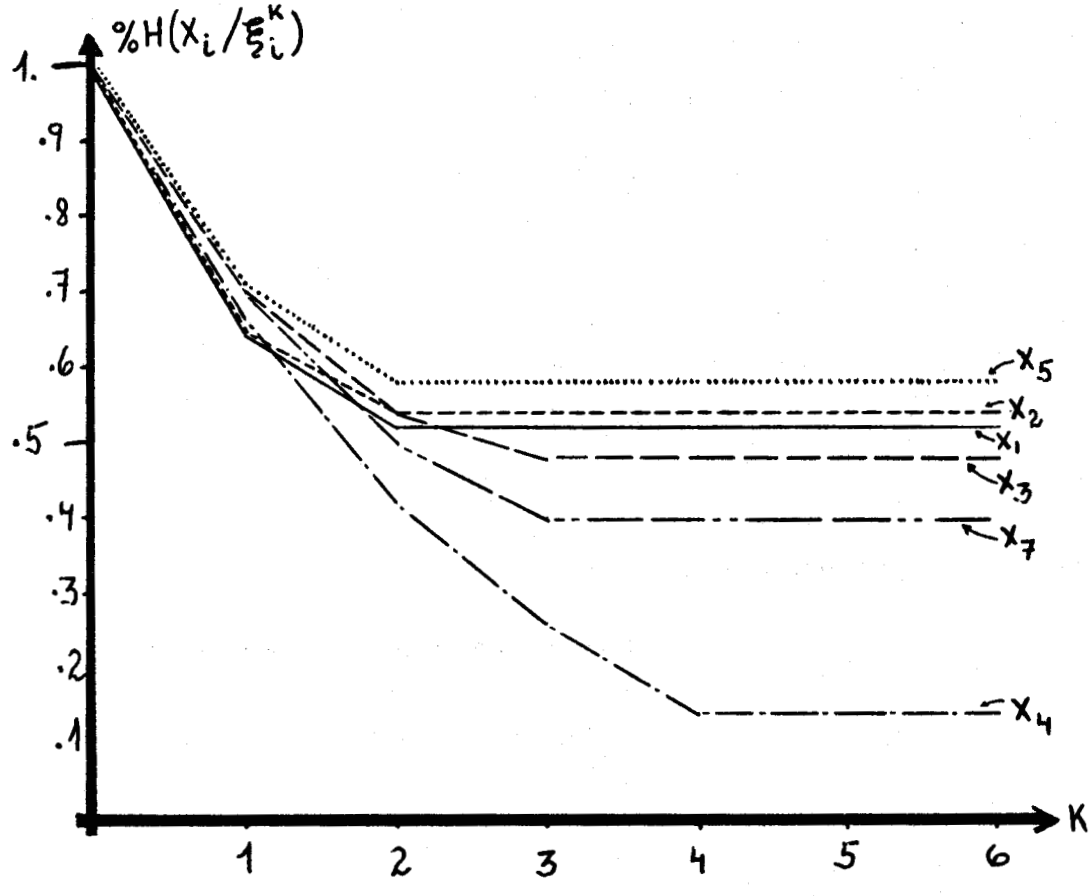


Figure IV.5 Pourcentage Non Expliqué de X_i sachant ξ_i^k

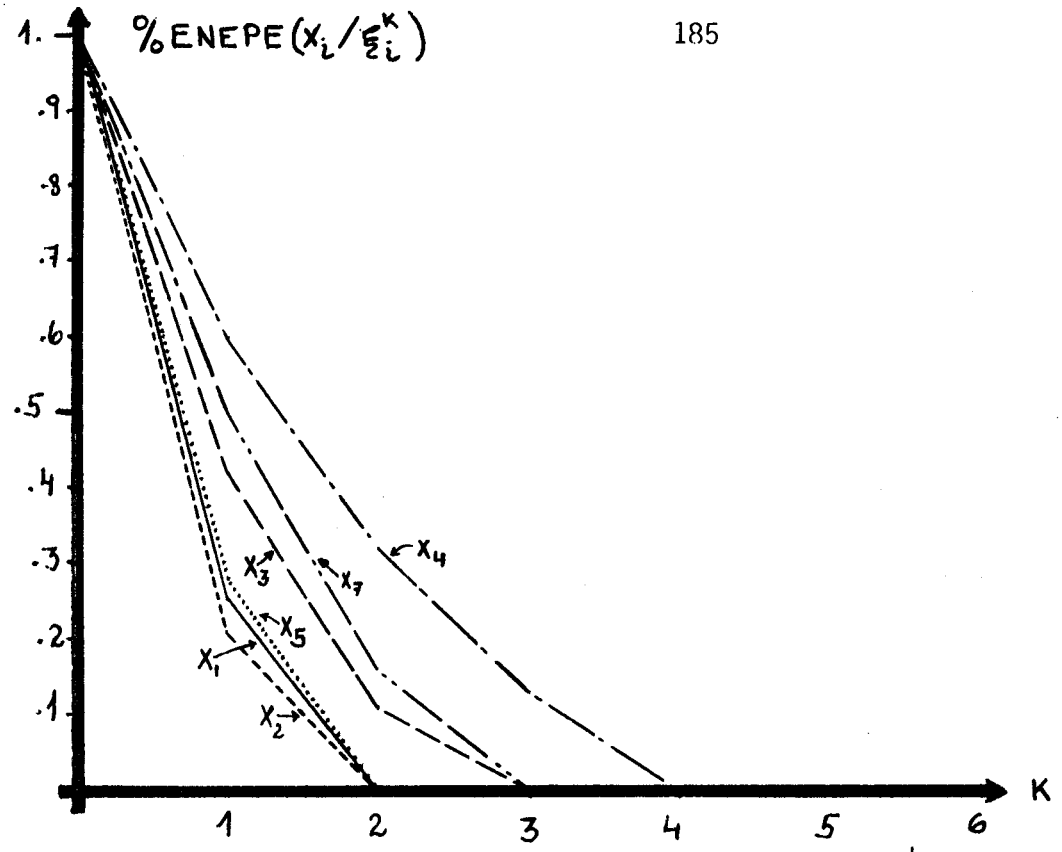


Figure IV.6 Pourcentage Encore Non Expliqué par ξ_i^k de la Partie Explicable de X_i

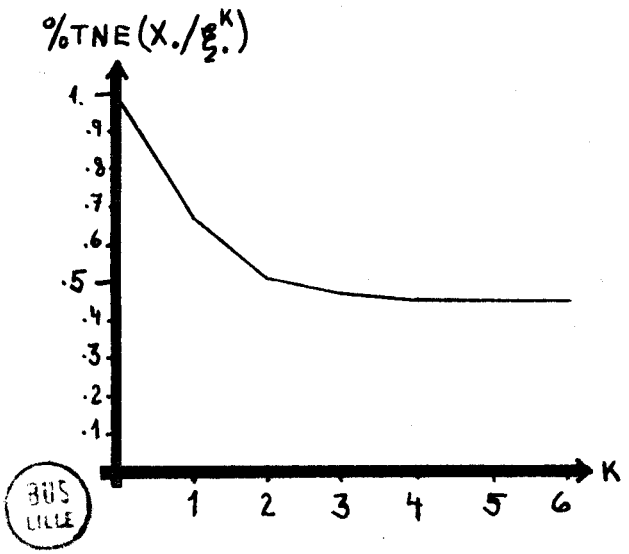


Figure IV.7 Pourcentage Total Non Expliqué sachant ξ_p^k

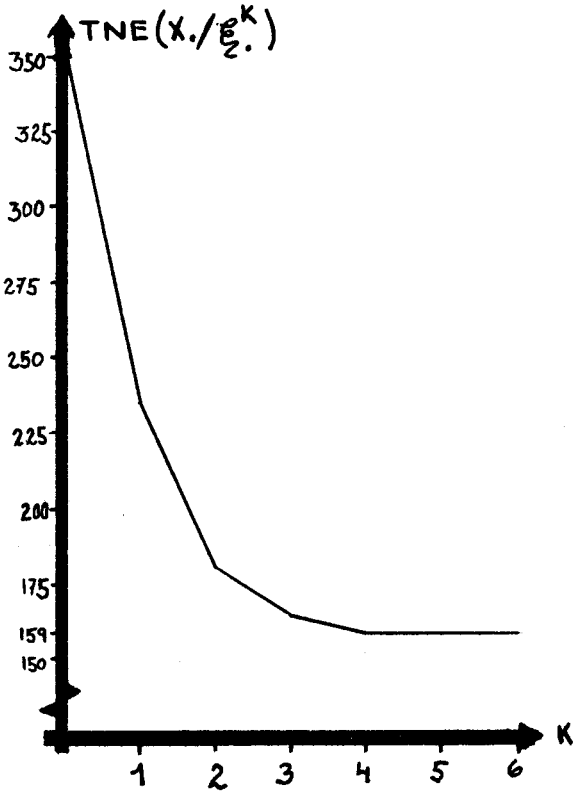


Figure IV.8 Total Non Expliqué sachant $\xi_.^k$

5 - ALGORITHME DE RECHERCHE EXHAUSTIVE (AE)

Nous allons maintenant étudier quatre algorithmes conduisant à déterminer (ou à approximer) les ξ^K, K=1,2,...,. Pour avoir une idée de la complexité de cette tâche, nous étudions tout d'abord un Algorithme de recherche Exhaustive (AE).

5.1 - ALGORITHME AE

On suppose $\xi := \{X_1, X_2, \ldots X_e\}$, et $\mathcal{O}_i$, $1 \leq i \leq e$, donnés. L'algorithme utilise deux vecteurs :

- $S_1, S_2, \ldots, S_{N-1}$, où S_K contient l'ensemble à K éléments qui jusqu'à présent explique le mieux X_i,
- $HS_1, HS_2, \ldots, HS_{N-1}$, où HS_K contient $H(X_i/S_K)$.

début ;

```
i ← 1;
Tant que i ≤ e faire ;
    S_K ← ∅, HS_K ← + ∞, pour K=1,2,...,|O_i|;
    Pour tout S ⊂ O_i faire;
        K ← |S|;
        Si H(X_i/S) < HS_K
            alors;
                S_K ← S;
                HS_K ← H(X_i/S);
        fpour;
    ftq;
    Ecrire S_K,HS_K, K=1,2,...,|O_i| ;
    i ← i+1;
ftq;
fin;
```

5.2 - OPTIMALITE DE L'ALGORITHME AE

Cet algorithme est évidemment optimal , étant donné qu'il parcourt tous les sous-ensembles possibles à la recherche de ceux qui minimisent le critère.

5.3 - ORDRE DE L'ALGORITHME AE

a) Cas statique

Le cas qui entraîne le plus de calculs est bien entendu celui où $\mathcal{E}^E = \{X_1^E, X_2^E, \ldots, X_N^E\}$ et $\mathcal{O}_i^E = \Sigma^E \setminus \{X_i^E\}$, donc $|\mathcal{E}^E| = N$ et $|\mathcal{O}_i| = N-1$.

Pour chacun des $X_i^E \in \mathcal{E}$ il y a $2^{N-1} - 1$ sous-ensembles de $\mathcal{O}_i$ à examiner, soit au total

$$\text{ORDRE (AE statique)} = N. (2^{N-1} - 1) \qquad [IV.25]$$

Dans l'Annexe 6, on calcule l'ordre ci-dessus, pour $N \leq 20$. Quant à la limite d'application, si l'on suppose qu'un ordinateur nécessite 1/1000 de seconde pour l'examen de chaque sous-ensemble , alors l'algorithme peut être appliqué jusqu'à N = 19. Dans ce cas le nombre de sous-ensembles à examiner est 9'961.453, tâche qui demande 2 h 46 mn.

b) Cas dynamique

Nous devons distinguer ici deux cas :

i) Problème d'Estimation

Il importe de savoir quelles variables (et avec quel décalage!) peuvent servir à expliquer une autre variable.

Considérons donc

$\xi^D = \{X_1(t), X_2(t),\ldots,X_e(t)\}$, ($\{X_{e+1},\ldots,X_N\}$ étant les variables d'entrée)
et

$\sigma_i^D = \Sigma \setminus \{X_1(t),\ldots,X_N(t)\}$

$= \{X_1(t-1),\ldots,X_1(t-(p-1)), X_2(t-1),\ldots,X_2(t-(p-1)),\ldots,X_N(t-(p-1))\}$

Nous avons donc $|\xi^D| = e$, $|\sigma_i^D| = N.(p-1)$. Pour chacune des e variables de ξ^D il faut examiner $2^{N.(p-1)}-1$ sous-ensembles de σ_i^D , soit au total

$$\text{ORDRE (AE dyn.Estim.)} = e.(2^{N.(p-1)}-1).$$

ii) Problème d'Analyse Structurale

Dans ce cas, il importe seulement de savoir quelles variables peuvent servir à expliquer une autre. Il suffit de considérer alors

$\xi^D = \{X_1(t),X_2(t),\ldots,X_e(t)\}$ ($\{X_{e+1},\ldots,X_N\}$ étant les variables d'entrée)

et

$\sigma_i^D = \{(X_1(t-1),\ldots,X_1(t-(p-1))),\ldots,(X_N(t-1),\ldots,X_N(t-(p-1)))\}$

Nous avons donc $|\xi^D| = e$, $|\sigma_i^D| = N$. On obtient alors

$$\text{ORDRE (AE dyn.Struct.)} = e.(2^N-1).$$

6 - ALGORITHME AGREGATIF (AA)

Pour chaque variable $X_i \in E$, cet algorithme AA construit une chaîne de sous-ensembles explicatifs $S_1, S_2, \ldots, S_{|\mathcal{O}_i|}$, avec $S_{K+1} = S_K \cup \{X_\alpha\}$, la variable X_α étant choisie de façon à rendre $H(X_i/S_K \cup \{X_\alpha\})$ aussi petit que possible.

6.1 - ALGORITHME AA

début;

$i \leftarrow 1$;

tant que $i \leq e$ faire;

  trouver la variable $X_\alpha \in \mathcal{O}_i$ telle que $H(X_i/X_\alpha) \leq H(X_i/X)$, $\forall X \in \mathcal{O}_i$;

  $S_1 \leftarrow \{X_\alpha\}$;

  $K \leftarrow 2$;

  tant que $K \leq |\mathcal{O}_i|$ faire;

    Trouver la variable $X_\alpha \in \mathcal{O}_i \setminus S_{K-1}$ telle que $H(X_i/S_{K-1}, X_\alpha) \leq H(X_i/S_{K-1}, X)$, $\forall X \in \mathcal{O}_i \setminus S_{K-1}$;

    $S_K \leftarrow S_{K-1} \cup \{X_\alpha\}$;

    $K \leftarrow K+1$;

  ftq;

  Ecrire S_K, $H(X_i/S_K)$, pour $K = 1,2,\ldots, |\mathcal{O}_i|$;

  $i \leftarrow i+1$;

ftq;

fin;

6.2 - OPTIMALITE DE L'ALGORITHME AA

Comme nous l'avons remarqué au paragraphe 3.1, les sous-ensembles explicatifs optimaux $\xi_i^1, \xi_i^2, \ldots, \xi_i^{|\mathcal{O}_i|}$ ne constituent pas forcément une chaine ordonnée par la relation $\subseteq$. En conséquence, tout algorithme dont le résultat est une chaîne de sous-ensembles peut ne pas trouver tous les optimums locaux.

Avec l'algorithme AA nous pouvons seulement être sûrs de trouver ξ_i^1 (car toutes les alternatives auront été examinées) et $\xi_i^{|\mathcal{O}_i|}$ (trivialement).

6.3 - ORDRE DE L'ALGORITHME AA

a) Cas statique

Soit $\mathcal{E}^E = \{X_1^E, X_2^E, \ldots, X_N^E\}$ et $\mathcal{O}_i^E = \Sigma^E \setminus \{X_i^E\}$.

Pour chacune des N variables de E, on peut dire :

. Déterminer S_1 requiert l'examen de N-1 alternatives,

. S_1 étant connu, déterminer S_2 requiert l'examen de N-2 alternatives,

⋮

. S_{K-1} étant connu, déterminer S_K requiert l'examen de N-K alternatives,

⋮

Ainsi donc, on obtient

$$\text{ORDRE (AA Stat.)} = N. \sum_{K=1}^{N-1} N-K = N \sum_{j=1}^{N-1} j = \frac{N^2(N-1)}{2} = \frac{N^3 - N^2}{2} \qquad [IV.26]$$

Dans l'Annexe 6, on calcule la fonction ci-dessus pour $N \leqslant 20$. Toujours avec l'hypothèse de 1/1000 de seconde pour l'examen de chaque alternative, cet algorithme pourrait être utilisé jusqu'à N = 262 variables. Dans un tel cas, le nombre d'alternatives est de 8'958.042, qui demandent 2h + 29 mn de temps de calcul.

b) Cas dynamique

Avec les mêmes ensembles $\mathcal{E}^D$ et σ_i^D décrits au paragraphe 5.3.b, on obtient facilement

i) Problème d'Estimation

$$\text{ORDRE (AA dyn.Estim.)} = e.\sum_{K=0}^{N.(p-1)} N.(p-1)-K = \frac{e.N(p-1).(N(p-1)+1)}{2}$$

ii) Problème d'Analyse Structurale

$$\text{ORDRE (AA dyn.Struc.)} = e.\sum_{K=0}^{N} N-K = \frac{e.N\ (N+1)}{2}$$

7 - ALGORITHME DÉSAGRÉGATIF (AD)

Cet algorithme est en quelque sorte l'inverse de celui qu'on vient de décrire. Dans le cas présent, pour chaque variable $X_i \in E$, cet algorithme AD construit une chaine $S_{|\sigma_i|}, S_{|\sigma_i|-1}, \ldots, S_2, S_1$, avec $S_{K-1} = S_K \setminus \{X_\alpha\}$, la variable X_α étant choisie de façon à rendre $H(X_i/S_{K-1})$ aussi petit que possible.

7.1 - ALGORITHME AD

```
début;
i ← 1;
Tant que i ≤ e faire;
    S_e ← σ_i;
    K ← |σ_i| - 1;
    Tant que K ≥ 1 faire;
        Trouver la variable X_α ∈ S_{K+1} telle que H(X_i/S_{K+1}\{X_α}) ≤ H(X_i/S_{K+1}\{X})
        ∀ X ∈ S_{K+1};
        S_K ← S_{K+1}\{X_α};
        K ← K-1;
    ftq;
    Ecrire S_K, H(X_i/S_K), pour K = 1,2,...,|σ_i|;
    i ← i+1;
ftq;
fin;
```

7.2 - OPTIMALITE DE L'ALGORITHME AD

Par la même raison que nous avons signalée au paragraphe 6.2, il n'est pas sûr que l'algorithme AD trouve tous les optimums locaux. Nous pouvons seulement être certains de trouver $\xi_i^{|\sigma_i|-1}$ (car toutes les alternatives auront été examinées) et $\xi_i^{|\sigma_i|}$ (trivialement).

7.3 - ORDRE DE L'ALGORITHME AD

a) Cas Statique

Soit $\mathcal{E}^E := \{X_1^E, X_2^E, \ldots, X_N^E\}$ et $\mathcal{O}_i^E = \Sigma^E \setminus \{X_i^E\}$. Comme pour l'algorithme AA, on peut dire que pour chacune des N variables de $\mathcal{E}^E$:

. Déterminer S_{N-1} requiert l'examen de N-1 alternatives,

. S_K étant connu, déterminer S_{K-1} requiert l'examen de K alternatives.

Au total donc, on obtient

$$\text{ORDRE (AD Stat.)} = N. \sum_{K=N-1}^{1} K = \frac{N^2(N-1)}{2} = \frac{N^3 - N^2}{2} \qquad [IV.27]$$

Ce résultat est identique à celui de AA Stat., ce qui veut dire que cet algorithme pourrait être appliqué jusqu'à N = 262 variables, nécessitant dans ce cas 2h + 29 min de calcul.

b) Cas dynamique

Avec les mêmes ensembles $\mathcal{E}^D$ et $\mathcal{O}_i^D$ décrits au paragraphe 5.3.b. on obtient :

i) Problème d'Estimation

$$\text{ORDRE (AD dyn. Estim)} = \text{ORDRE (AA dyn.Estim)} = \frac{e.N.(p-1)(N(p-1)+1)}{2}$$

ii) Problème d'Analyse Structurale

$$\text{ORDRE (AD dyn.Struc.)} = \text{ORDRE (AA dyn.Struc.)} = \frac{e.N.(N+1)}{2}$$

8 - ALGORITHME D'APPROXIMATIONS SUCCESSIVES (AAS)

8.1 - OPERATEURS CONTRACTANTS

Pour chaque $X_i \in \mathcal{X}$ on peut définir :

$$\mathcal{P}(\mathcal{O}_i) = \{S : S \subseteq \mathcal{O}_i\}$$

$$\mathcal{P}_K(\mathcal{O}_i) = \{ S : S \subseteq \mathcal{O}_i,\ |S| = K \}\ ,\ K=0,1,2,\ldots,|\mathcal{O}_i|\ .$$

Les algorithmes AA et AD que nous venons d'étudier se résument à une application répétée pour chaque $X_i \in \mathcal{X}$ des opérateurs suivants :

$$\text{(Ajouter:)}\quad A:\mathcal{P}_K(\mathcal{O}_i) \longrightarrow \mathcal{P}_{K+1}(\mathcal{O}_i)\ ,\ 0 \leq K \leq |\mathcal{O}_i| - 1 \qquad [IV.28]$$

$$S \rightsquigarrow A(S) = S \cup \{X_\alpha\}$$

où $X_\alpha \in \mathcal{O}_i \setminus S$, et $\forall X \in \mathcal{O}_i \setminus S : H(X_i/S\cup\{X_\alpha\}) \leq H(X_i/S\cup\{X\})$.

$$\text{(Retirer)}\quad R:\mathcal{P}_K(\mathcal{O}_i) \longrightarrow \mathcal{P}_{K-1}(\mathcal{O}_i),\ 1 \leq K \leq |O_i|\ , \qquad [IV.29]$$

$$S \rightsquigarrow R(S) = S \setminus \{X_\alpha\}$$

où $X_\alpha \in S$ et $\forall X \in S : H(X_i/S\setminus\{X_\alpha\}) \leq H(X_i/S\setminus\{X\})$.

A l'aide des opérateurs A et R, nous pouvons définir les deux opérateurs suivants, pour $K=1,2,\ldots,|\mathcal{O}_i| - 1$:

$$AoR : \mathcal{P}_K(\mathcal{O}_i) \longrightarrow \mathcal{P}_K(\mathcal{O}_i) \qquad [IV.30]$$

$$S \mapsto AoR(S) := A(R(S))$$

$$RoA : \mathcal{P}_k(\mathcal{O}_i) \longrightarrow \mathcal{P}_K(\mathcal{O}_i) \qquad [IV.31]$$

$$S \mapsto RoA(S) := R(A(S))$$

Proposition IV.8 (SV)

Quel que soit $X_i \in \mathcal{E}$, et quel que soit K, $1 \leqslant K \leqslant |\mathcal{O}_i| - 1$, on a :

$\forall S \in \mathcal{P}_K(\mathcal{O}_i)$:

$$H(X_i/S) \geqslant H(X_i/AoR(S)), \text{ et}$$
$$H(X_i/S) \geqslant H(X_i/RoA(S)).$$

Démonstration :

Soit $X_i \in \mathcal{E}$ et $S \in \mathcal{P}_K(\mathcal{O}_i)$, pour un K : $1 \leqslant K \leqslant |\mathcal{O}_i| - 1$.

$H(X_i/R(S)) = \text{Min } \{H(X_i/S\backslash\{X\}) : X \in S\} = H(X_i/S\backslash\{X_\beta\})$,

pour un certain $X_\beta \in S$. D'autre part, nous avons :

$$H(X_i/AoR(S)) := H(X_i/A(R(S))) = H(X_i/A(S\backslash\{X_\beta\}))$$
$$= \text{Min } \{H(X_i/(S\backslash\{X_\beta\}) \cup \{X\}) : X \in \mathcal{O}_i\backslash(S\backslash\{X_\beta\})\}$$

et puisque $X_\beta \in \mathcal{O}_i\backslash(S\backslash\{X_\beta\})$, alors

$$\leqslant H(X_i/(S\backslash\{X_\beta\}) \cup \{X_\beta\})$$
$$= H(X_i/S).$$

La démonstration pour RoA(S) est tout à fait similaire à celle-ci.

(F.D.)

La conséquence du caractère contractant de AoR et RoA est immédiate :

Proposition IV.9 (SV)

Tous les sous-ensembles explicatifs optimaux ξ_i^K sont des points fixes des opérateurs AoR et RoA, c'est-à-dire,

$\forall X_i \in \mathcal{E}$, $\forall K$: $1 \leqslant K \leqslant |\mathcal{O}_i| - 1$:

$$H(X_i/\xi_i^K) = H(X_i/AoR(\xi_i^K)) = H(X_i/RoA(\xi_i^K)).$$

Démonstration :

D'une part, la proposition IV.8 (SV) implique

$$H(X_i/AoR(\xi_i^K)) \leqslant H(X_i/\xi_i^K) \geqslant H(X_i/RoA(\xi_i^K)).$$

D'autre part, étant donné que $AoR(\xi_i^K)$ et $RoA(\xi_i^K)$ appartiennent à $\mathcal{P}_K(O_i)$, par définition de ξ_i^K on a :

$$H(X_i/AoR(\xi_i^K)) \geqslant H(X_i/\xi_i^K) \leqslant H(X_i/R.A(\xi_i^K)).$$

(F.D.)

REMARQUE

Il convient de signaler que AoR $\neq$ RoA, comme le montre l'exemple suivant :

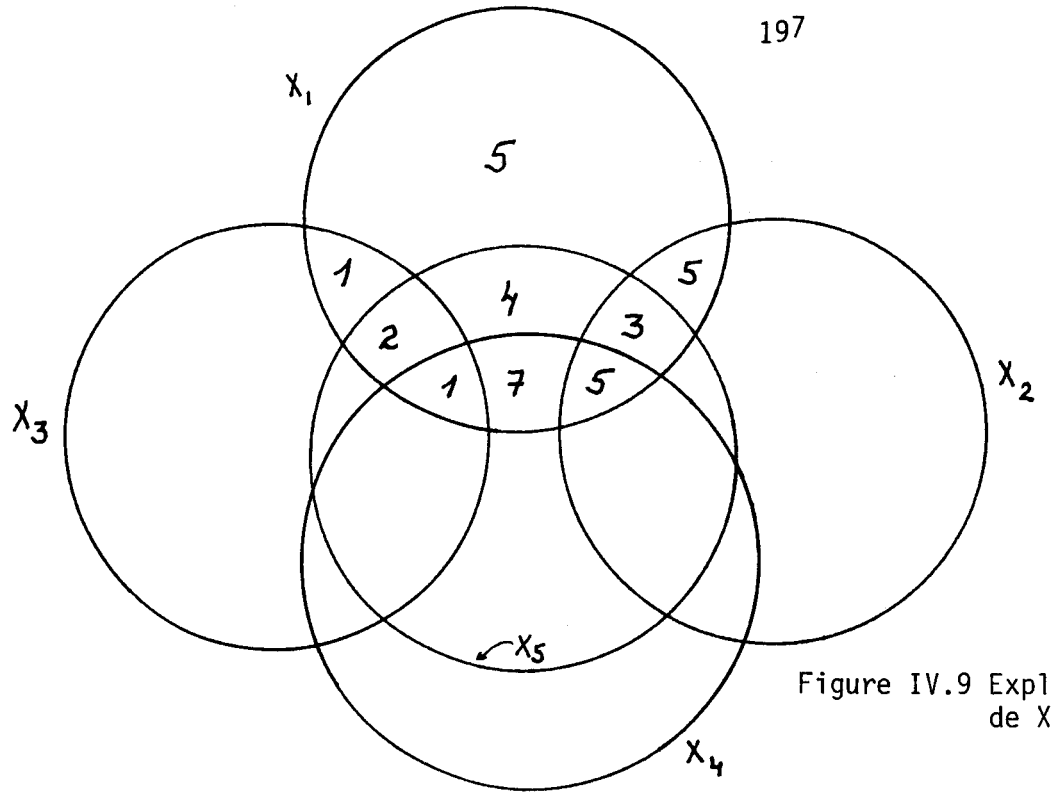


Figure IV.9 Explication de X_1

Supposons qu'il s'agit de trouver un ensemble explicatif de X_1 à deux variables. Soit $S = \{X_3, X_4\}$.

$S = \{X_3, X_4\}$ $\qquad H(X_1/S) = 5+5+4+3 = 17$

on calcule facilement :

$A(S) = \{X_2, X_3, X_4\}$ $\qquad H(X_1/A(S)) = 5+4 = 9$

$R(A(S)) = R(\{X_2, X_3, X_4\}) = \{X_2, X_4\}$ $\qquad H(X_1/R(A(S))) = 5+4+1+2 = 12$

D'autre part

$R(S) = \{X_4\}$ $\qquad H(X_1/R(S)) = 5+5+3+4+2+1 = 20$

$A(R(S)) = \{X_4, X_5\}$ $\qquad H(X_1/A(R(S))) = 5+5+1 = 11$

8.2 - ALGORITHME AAS

On dira qu'un sous-ensemble $S \subseteq \mathcal{P}(\mathcal{O}_i)$ est *stable* ssi c'est un point fixe aussi bien de AoR que de RoA. On vient de voir par ailleurs que tout sous-ensemble explicatif Optimal ξ_i^K est stable. Cependant, il ne s'agit que d'une condition nécessaire d'optimalité, car il peut y avoir des sous-ensembles stables qui ne soient pas optimaux.

L'algorithme d'Approximations successives que nous proposons cherche pour chaque $X_i \in \mathcal{E}$ et pour chaque K, $1 \leq K \leq |\mathcal{O}_i| - 1$, un sous-ensemble de $\mathcal{O}_i$ vérifiant la condition nécessaire d'optimalité. Pour cela, pour chaque K on se donne un sous-ensemble à K éléments de $\mathcal{O}_i$, quelconque, et on lui applique les opérateurs AoR et RoA autant de fois que cela est possible, jusqu'à stabilisation. La convergence est assurée par le caractère contractant des opérateurs.

Plus concrêtement, les boucles principales des algorithmes AA et AD deviennent des sous-programmes, appellés par l'algorithme AAS :

```
début;
i ← 1;
Tant que i ≤ e faire;
    K ← 1;
    Tant que K ≤ |Oi| - 1 faire;
        Générer (ou lire) un élément S de PK(Oi);
        Stable ← "faux";
        Tant que ¬Stable faire;
            Tant que AoR(S) ≠ S faire;
                S ← AoR(S);
            ftq;
            Tant que RoA(S)≠ S faire;
                S ← RoA(S);
            ftq;
            Si AoR(S) = RoA(S) = S
               alors;
                    Stable ← "vrai";
            fsi;
        ftq;
```

```
        Ecrire S, H(X_i/S);
        K ← K+1;
    ftq;
    i ← i+1;
ftq;
fin;
```

Rien n'empêche, bien entendu, que pour chaque $X_i \in \mathcal{E}$ et pour chaque K, $1 \leq K \leq |\mathcal{O}_i| - 1$, on se donne non pas un mais plusieurs sous-ensembles initiaux tirés au hasard, pour les "stabiliser". Cela augmentera sans doute les "chances" de trouver les ξ_i^K .

8.3 - OPTIMALITE DE L'ALGORITHME AAS

Cet algorithme fournit pour chaque $X_i \in \mathcal{E}$ et pour chaque K, $1 \leq K \leq |\mathcal{O}_i| - 1$, un (ou plusieurs) sous-ensembles stables. Bien que ceux-ci ne forment pas forcément une chaîne ordonnée, nous pouvons uniquement dire qu'ils vérifient une condition nécessaire d'optimalité. Par ailleurs étant donné que AAS réunit les propriétés de AA et AD, nous pouvons être sûrs de trouver ξ_i^1 et $\xi_i^{|\mathcal{O}_i|-1}$, ainsi que $\xi_i^{|\mathcal{O}_i|}$ (ce dernier étant trivial).

3.4 - ORDRE DE L'ALGORITHME AAS

a) Cas Statique

Soit $\mathcal{E}^E = \{X_1^E, X_2^E, \ldots, X_N^E\}$ et $\mathcal{O}_i^E = \Sigma^E \setminus \{X_i^E\}$. Voyons d'abord le nombre d'alternatives examinées lors de l'application de l'opérateur AoR à un $S \subseteq \mathcal{O}_i$, $|S| = K$:

le calcul de R(S) requiert l'examen de K alternatives (lequel des K éléments de S retirer ?). On obtient comme résultat un ensemble de K-1 éléments. Ensuite, l'application de l'opérateur A à R(S) requiert l'examen de N-1 - (K-1) = N-K alternatives (lequel des (N-1) - (K-1) élément de $(\Sigma^E \setminus \{X_i^E\}) \setminus R(S)$ convient-il d'ajouter à R(S)?). Ainsi, une application de AoR à un ensemble de K élément requiert l'examen de K + (N-K) = N alternatives (quel que soit X_i et quel que soit K). On peut montrer facilement qu'il en est de même pour RoA.

D'autre part, si pour chacune des N variables de E on se donne α solutions initiales tirées au hasard, et que pour chacune d'elles on applique β fois un opérateur contractant, on obtient

$$\text{ORDRE (AAS stat.)} = \alpha.\beta.N \sum_{K=1}^{N-1} N = \alpha\ \beta\ (N-1).N^2 = \alpha\ \beta.(N^3 - N^2) \qquad [IV.32]$$

Le calcul de cette fonction pour $N \leqslant 20$ est présenté dans l'Annexe 6. Enfin, avec l'hypothèse de 1/1000 de seconde par alternative examinée, et en prenant $\alpha.\beta = 18$ (ce qui correspond à 3 solutions initiales auxquelles on applique 6 fois un opérateur contractant, ...),cet algorithme pourrait être utilisé jusqu'à N=80 variables. Dans ce cas le nombre d'alternatives est de 9'100.800 qui nécessitent 2h + 31 min de temps de calcul.

b) Cas Dynamique

Avec les ensembles E^D et σ_i^D décrits au paragraphe 5.3.6., on applique la même méthode de comptage qui vient d'être utilisée. Les résultats sont les suivants :

i) Problème d'Estimation

$$\text{ORDRE (AAS Dyn.Estim)} = \alpha.\beta.e.\sum_{K=1}^{N(p-1)-1} N(p-1) = \alpha.\beta.e.N(p-1).(N(p-1)-1).$$

ii) Problème d'Analyse Structurale

$$\text{ORDRE (AAS Dym.Struc.)} = \alpha.\beta.e \sum_{K=1}^{N-1} N = \alpha.\beta.e.N(N-1).$$

9 - APPLICATION A L'ANALYSE STRUCTURALE

Plaçons nous maintenant dans le cas où l'intérêt principal est de déterminer la structure des liaisons entre les variables du système. Prenons donc $\mathcal{E}$ et $\mathcal{O}_i$ définis ainsi :

* *Système Statique*

$$\mathcal{E}^E = \{X_1^E, X_2^E, \ldots, X_N^E\} = \Sigma^E ;$$

pour i = 1,2,... N: $\mathcal{O}_i^E = \Sigma^E \setminus \{X_i^E\}$

c'est-à-dire, on suppose que chacune des variables du système peut à la limite être influencée par toutes les autres variables du système.

* *Système Dynamique*

$$\mathcal{E}^D = \{X_1(t), X_2(t), \ldots, X_e(t)\} \text{ (où } X_{e+1}, X_{e+2}, \ldots, X_N \text{ sont les variables d'entrée);}$$

pour i = 1,2,...,e : $\mathcal{O}_i^D = \Sigma \setminus \{X_1(t), \ldots, X_N(t)\}$

$$= \{(X_1(t-1), \ldots, X_1(t-(p-1))), \ldots, (X_N(t-1), \ldots, X_N(t-(p-1)))\}$$

ainsi, chacune des variables (sauf celles d'entrée) à l'instant t, peut être influencée par toutes les variables du système aux instants précédents.

9.1 - GRAPHE DE DEPENDANCE

Avec $\mathcal{E}$ et $\mathcal{O}_i$ définis précédemment, et les sous-ensembles ξ_i^K déterminés (ou approximés) au moyen des algorithmes que nous avons décrits, on peut construite les *Graphes de Dépendance* (ou d'influence) ainsi :

$$G_K = (\Sigma, U_K), \quad K = 1,2,\ldots,N$$

avec $\Sigma = \{X_1, X_2, \ldots, X_N\}$, l'ensemble de sommets;

$U_K = \{(X_i, X_j) : X_i \in \xi_j^K\}$, l'ensemble d'arcs.

Autrement dit, c'est un graphe où les sommets sont les variables du système, et où un arc (X_i, X_j) peut être interprété comme : "la variable X_i contribue, avec d'autres, à déterminer ou à expliquer la variable X_j".

Ce recours aux graphes pour exprimer l'existence de liens entre les variables, est assez classique dans la littérature [CARP 80a, 80b; DELA 79; GABA 78; NEPO 78, RICH 75, chap.II; STEW 65, parmi d'autres]. De ce fait, il existe de nombreuses techniques pour la décomposition, la hiérarchisation ou la visualisation des graphes.

Mais la caractéristique la plus importante à notre avis des graphes G_K que nous proposons est leur caractère multi-dimensionnel. En effet, dans les graphes couramment utilisés, un arc (X_i, X_j) représente l'existence d'une liaison entre X_i et X_j mesurée avec un critère *binaire*, de la forme

$$(X_i, X_j) \in U \Leftrightarrow S(X_i, X_j) \geqslant \text{seuil donné},$$

S étant un indice de similarité quelconque (par exemple le coefficient de corrélation linéaire, ou bien la transinformation $\overleftrightarrow{I}(X_i : X_j)$, etc .).

Or dans notre cas, un arc (X_i, X_j) indique que X_i est liée à X_j, mais surtout, que cette liaison se passe dans le cadre d'une influence *conjointe* avec les autres variables de ξ_j^K. On a déjà montré (c.f.chapII,§.9.3) comment d'importants aspects de la structure d'un système ne sont décelables qu'avec une approche "multidimensionnelle", et qu'ils passent tout à fait inaperçus lors de n'importe quelle analyse binaire.

9.2 - DECOMPOSITION, HIERARCHISATION ET VISUALISATION DE GRAPHES

L'analyse des graphes étant un sujet largement étudié et publié, il ne nous apparait pas nécessaire de faire leur présentation dans ce mémoire. Nous voulons tout de même mentionner quelques travaux particulièrement intéressants.

Marc RICHETIN a présenté dans le chapitre II de sa thèse [RICH 75] quelques méthodes de décomposition et de hiérarchisation de graphes qui aident à la mise en oeuvre efficace des techniques de Commande Hiérarchisée. D'autre part, D.GABAY, P.NEPOMIASTCHY proposent des algorithmes qui permettent de transformer le graphe d'un système sous des formes qui simplifient sa résolution et son optimisation [GABA 78; NEPO 78]. Mentionnons aussi les recherches récentes de M.J.CARPANO et M.DELARCHE, qui ont développé des méthodes informatiques pour simplifier la représentation et la visualisation de graphes de grande taille [CARP 80a, 80b; DELA 79]. Enfin, J.N.WARFIELD a fait un travail considérable d'analyse structurale de systèmes socio-économiques à l'aide de la Théorie des Graphes [WARF 73a, 73b, 74a, 74b, 74c, 76,77].

10 - APPLICATION A LA RECONNAISSANCE DE FORMES

Les algorithmes que nous venons d'étudier sont d'une application immédiate au problème de la Reconnaissance de Formes (ou Classification Automatique) avec apprentissage. Le problème peut être énoncé ainsi :

On suppose que l'on dispose d'une *Population d'Apprentissage* $\Omega = \{\omega_1, \omega_2, \ldots, \omega_L\}$. Ces objets se classifient dans les *formes* (ou classes) $\varphi_1, \varphi_2, \ldots, \varphi_m$. Pour cela, on possède un ensemble de *Paramètres* $X_1, X_2, \ldots, X_N$ qui peuvent être mesurés sur chacun des objets. Le diagramme suivant résume cette formulation :

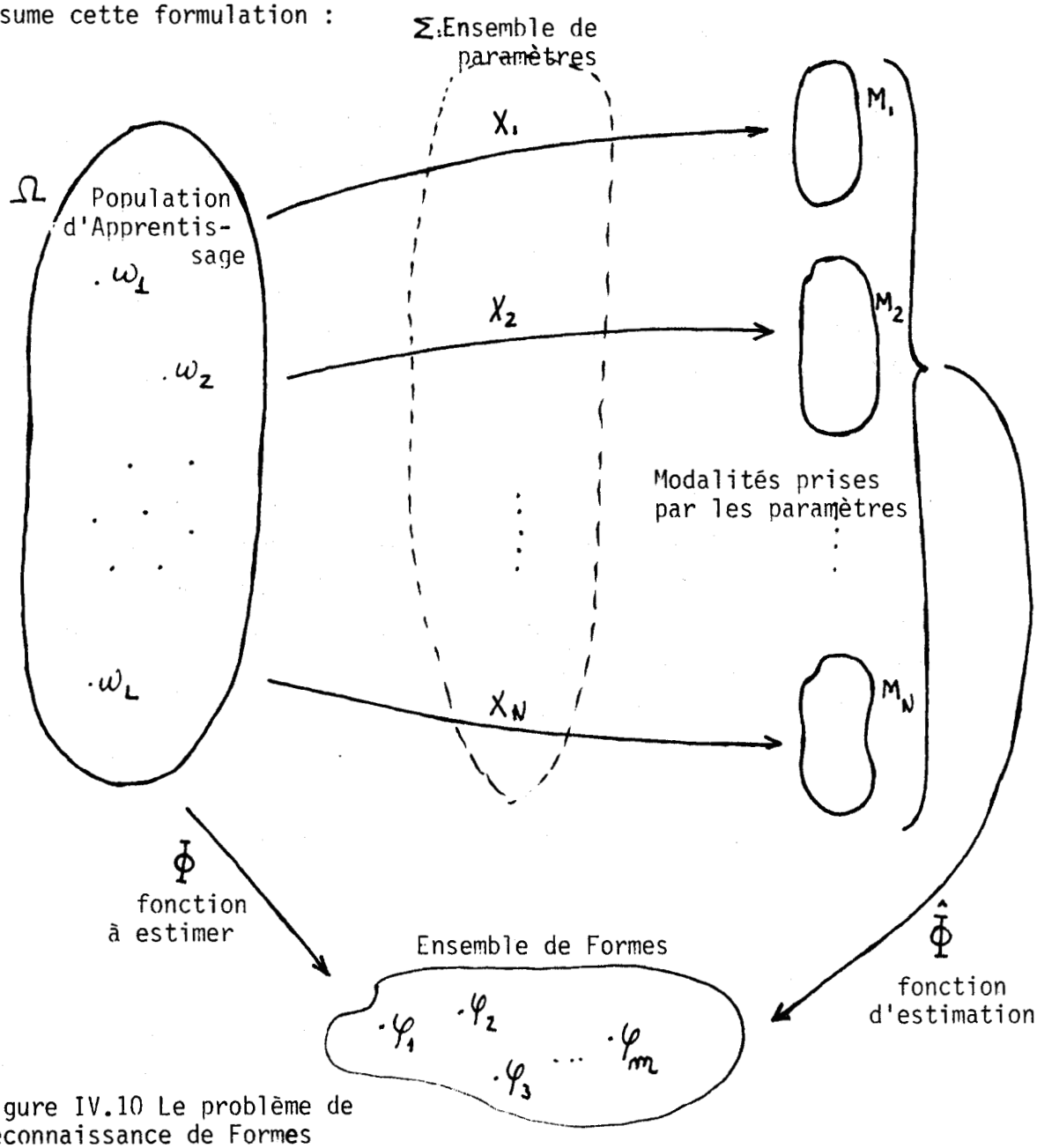


Figure IV.10 Le problème de Reconnaissance de Formes

L'objectif est alors d'estimer la fonction d'affectation Φ qui à chaque objet ω fait correspondre sa forme $\Phi(\omega)$, et cela à l'aide des paramétres $X_1,\ldots, X_N$.

Le problème exprimé dans la notation que nous avons utilisée est :

$$\Sigma = \{\Phi, X_1, X_2, \ldots, X_N\},$$

$$\mathcal{E} = \{\Phi\},$$

$$\mathcal{O}_\Phi = \{X_1, X_2, \ldots, X_N\}.$$

Il s'agit donc de :

i) trouver le sous-ensemble ξ de paramètres qui fournissent le plus d'information pour déterminer quelle est la forme ou classe à laquelle appartient chaque objet,

ii) une fois le sous-ensemble ξ trouvé, construire une fonction $\hat{\Phi}(\xi)$ qui approxime le mieux la fonction Φ (on suppose bien entendu que sur la population d'apprentissage Ω la fonction Φ est connue).

L'indice $H(\Phi/X_1,\ldots,X_N)$ (ou % $H(\Phi/X_1,\ldots,X_N)$) nous indique à quel point la classification peut être déduite des paramètres $X_1, X_2, \ldots, X_N$. Si cet indice a une faible valeur, on peut, à l'aide des algorithmes étudiés, trouver une suite $\xi^1, \xi^2, \ldots, \xi^N \subseteq \mathcal{O}_i$ de sous-ensembles permettant une estimation de Φ de plus en plus exacte. On pourra, sur la base des valeurs décroissantes $H(\Phi/\xi^K)$, $K = 1,\ldots,N$, faire un compromis entre l'exactitude voulue et le nombre de paramètres qu'il faudrait prendre en compte.

Une fois un ξ^{K_o} choisi, il reste à construire la fonction

$$\hat{\Phi}: M_{\xi^{k_o}} \longrightarrow \{\varphi_1, \varphi_2, \ldots, \varphi_m\}$$

(rappelons que $M_{\xi^{k_o}} := \prod_{i:X_i \in \xi^{k_o}} M_i$)

qui estime la forme à laquelle appartient un sujet quelconque en fonction de valeurs prises par les paramètres de l'ensemble ξ^{K_0}. Nous discuterons de la construction d'une telle fonction $\hat{\Phi}$ au paragraphe suivant.

II - SUR L'IDENTIFICATION DES FONCTIONS LIANT LES VARIABLES

Les algorithmes que nous avons étudiés dans ce chapitre permettent de trouver pour chaque variable $X_i \in \mathcal{E}$ l'ensemble de variables $\xi_i^{K_o} \subseteq \theta_i$ qui fournit le plus d'information sur X_i. Le cardinal K_o, comme nous l'avons dit, est fixé par un compromis entre l'exactitude requise et la dimension du modèle.

A ce stade subsiste alors le problème de trouver la fonction $f_i(\xi_i^{K_o})$ qui sert à approximer X_i, autrement dit, qui compte-tenu des modalités prises par les variables de $\xi_i^{K_o}$ donne une estimation de la modalité prise par X_i. Il est intéressant de savoir que quelle que soit cette fonction f_i, il y aura forcément une perte d'information.

II.1 - PERTE D'INFORMATION LORS DE TRANSFORMATIONS

Montrons tout d'abord les deux Lemmes suivants

Lemme IV.1

Quels que soient a,b,p,q tels que

$$0 \leqslant a, b \leqslant 1$$
$$0 < p, q \leqslant 1$$
$$a \leqslant p, b \leqslant q$$
$$a + b \leqslant p + q \leqslant 1$$

alors

$$a.\log \frac{a}{p} + b.\log \frac{b}{q} \geqslant (a+b).\log \frac{(a+b)}{(p+q)}$$

Démonstration :
==========

Il nous faut introduire les changements de notation suivants :

$M := a+b$ $\qquad$ $N := p+q$

$A := \frac{a}{a+b}$ $\qquad$ $C := \frac{p}{p+q}$

$B := \frac{b}{a+b}$ $\qquad$ $D := \frac{q}{p+q}$

on a alors

$a = A.M$ $\qquad$ $p = C.N$

$b = B.M$ $\qquad$ $q = D.N$

$A + B = 1$ $\qquad$ $C + D = 1$

avec cette nouvelle notation, nous allons montrer que la quantité suivante est positive :

$$a.\ \log \frac{a}{p} + b.\log \frac{b}{q} - (a+b).\log \frac{(a+b)}{(p+q)}$$

$$= AM.\log \frac{AM}{CN} + BM.\log \frac{BM}{DN} - M.\log \frac{M}{N}$$

$$= AM.\log AM - AM.\log CN + BM.\log BM - BM.\log DN - M.\log M + M.\text{Log } N$$

$$= AM.\log A + AM.\log M - AM.\log C - AM.\log N + BM.\log B + BM.\log.M$$

$$- BM.\ \log D - BM.\ \log N - M.\log M + M.\ \log N$$

$$= M.\ (A.\log A + B.\log B) + (A+B).M.\log M - M.\log M$$

$$- (A+B).M.\log N + M.\log N - M.\ (A.\log C + B.\log D)$$

étant donné que $A+B = 1$, on obtient :

$= M.(A.\log A + B.\log B) - M.(A.\log C + B.\log D)$

$= M.[(A.\log A + B.\log B) - (A.\log C + B.\log D)]$

Cette dernière quantité est positive ou nulle, d'une part parce que $M > 0$, et d'autre part parce que le terme entre [ ] est lui aussi positif ou nul. En effet A,B et C,D véfifient les conditions du Lemme II.1 (page 48).

(F.D.)

Ce Lemme IV.1 peut être généralisé par récurrence :

Lemme IV.2

Quels que soient $a_1, a_2, \ldots a_r$ et $b_1, b_2, \ldots, b_r$, avec

$$\forall i : 0 \leqslant a_i \leqslant 1;\ 0 < b_i \leqslant 1;\ a_i < b_i ;\ \sum_{i=1}^{r} a_i \leqslant \sum_{i=1}^{r} b_i \leqslant 1$$

alors

$$\sum_{i=1}^{r} a_i . \log \frac{a_i}{b_i} \geqslant (\sum_{i=1}^{r} a_i) \log \frac{(\sum_{i=1}^{r} a_i)}{(\sum_{i=1}^{r} b_i)}$$

Nous pouvons maintenant établir la proposition suivante :

Proposition IV.10

Pour H = Entropie, quelles que soient les variables X et S, et quelle que soit la fonction f définie sur l'ensemble des modalités de S, on a :

$$H(X/S) \leqslant H(X/f(S))$$

Interprétation :
==========

L'incertitude de X sachant déjà S ne peut qu'augmenter, quelle que soit la transformation que l'on fasse sur S.

Démonstration :
==========

Soit Z = f(S). Supposons que

X prends les valeurs $x_1,x_2,\ldots,x_q$, avec les probabilités $p(x_1),p(x_2),\ldots,p(x_q)$;

S prends les valeurs $s_1,s_2,\ldots,s_r$, avec les probabilités $p(s_1),p(s_2),\ldots,p(s_r)$;

Z prends les valeurs $z_1,z_2,\ldots,z_u$, avec les probabilités $p(z_1),p(z_2),\ldots,p(z_u)$;

On ne prend pas de généralité en supposant toutes ces probabilités strictement positives . Etant donné que Z = f(S), nous avons les relations :

$$p(z_K) = \sum_{j:f(s_j)=z_K} p(s_j)$$

$$p(x_i,z_K) = \sum_{j:f(s_j)=z_K} p(x_i,s_j), \text{ etc },$$

donc

$$p(x_i/z_K) := \frac{p(x_i,z_K)}{p(z_K)} = \sum_{j:f(s_j)=z_K} p(x_i,s_j) \Big/ \sum_{j:f(s_j)=z_K} p(s_j)$$

D'après l'affirmation [II.18] (page 79) nous avons :

$$- H(X/Z) = - \sum_{K=1}^{u} p(z_K).H(X/z_K)$$

$$= - \sum_{K=1}^{u} p(z_K).\left[- \sum_{i=1}^{q} p(x_i/z_K).\log\, p(x_i/z_K) \right]$$

$$= \sum_{K=1}^{u} \sum_{i=1}^{q} p(x_i,z_K).\log \frac{p(x_i,z_K)}{p(z_K)}$$

$$= \sum_{K=1}^{u} \sum_{i=1}^{q} \left[\left[\sum_{j:f(s_j)=z_K} p(x_i,s_j) \right] .\log \frac{\left[\sum_{j:f(s_j)=z_K} p(x_i,s_j) \right]}{\left[\sum_{j:f(s_j)=z_K} p(s_j) \right]} \right] \qquad [IV.33]$$

Pour chaque i,K, les $p(x_i,s_j)$ et $p(s_j)$ pour $j:f(s_j) = z_K$, vérifient les conditions du Lemme IV.2, c'est-à-dire

$$\forall_j : f(s_j) = z_K :$$

$$0 \leq p(x_i,s_j) \leq 1;\ 0 < p(s_j) \leq 1;\ p(x_i,s_j) \leq p(s_j);$$

$$\sum_j p(x_i,s_j) \leq \sum_j p(s_j) \leq 1$$

Ainsi donc:

$$[IV.33] \leq \sum_{K=1}^{u} \sum_{i=1}^{q} \left[\sum_{j:f(s_j)=z_K} p(x_i,s_j).\log \frac{p(x_i,s_j)}{p(s_j)} \right] \qquad (*)$$

$$= \sum_{K=1}^{u} \sum_{j:f(s_j)=z_K} \sum_{i=1}^{q} p(x_i,s_j).\log \frac{p(x_i,s_j)}{p(s_j)}$$

(*) Il est important de remarquer que si la fonction f est injective ($s_\alpha \neq s_\beta \Rightarrow f(s_\alpha) \neq f(s_\beta)$), alors la somme entre [ ] ne contient qu'un seul j pour chaque K,i. Dans ce cas on obtient une égalité.

$$= \sum_{j=1}^{r} \sum_{i=1}^{q} p(x_i, s_j) . \log \frac{p(x_i, s_j)}{p(s_j)}$$

$$= - \sum_{j=1}^{r} p(s_j) \left[- \sum_{i=1}^{q} p(X_i/s_j) . \log p(X_i/s_j) \right]$$

$$= - \sum_{j=1}^{r} p(s_j) \, H(X/s_j)$$

$$= - H(X/S)$$

Ainsi, $- H(X/Z) = - H(X/f(S)) \leqslant - H(X/S)$, c'est-à-dire

$$H(X/S) \leqslant H(X/f(S)).$$

(F.D.)

Proposition IV.11

Pour H = Entropie, on peut dire :

Quels que soient $A,B \in \mathcal{P}(\Sigma)$, et quelles que soient les fonctions f et g définies dans l'ensemble de modalités de A et B respectivement, on a

$$\overleftrightarrow{I}(A:B) \geqslant \overleftrightarrow{I}(f(A) : g(B))$$

Interprétation :

Quelles que soient les transformations (univoques) que l'on fasse sur deux ensembles de variables, on ne peut que réduire leur échange d'information.

Démonstration :

$$\begin{aligned}
\overleftrightarrow{I}(A:B) &= H(A) - H(A/B) \\
&\geqslant H(A) - H(A/g(B)) \\
&= \overleftrightarrow{I}(A:g(B)) \\
&= H(g(B)) - H(g(B)/A) \\
&\geqslant H(g(B)) - H(g(B)/f(A)) \\
&= \overleftrightarrow{I}(f(A):g(B))
\end{aligned}$$

(F.D.)

11.2 - LE CONCEPT D'ESTIMATION OPTIMALE

Si l'on a déjà déterminé l'ensemble S des variables qui donnent le plus d'information sur une certaine variable X, le choix de la

fonction f telle que f(S) soit une "bonne estimation" de X, dépend aussi bien de la nature algébrique des variables, que du critère choisi pour mesurer ce qui est une "bonne estimation". Soit :

$$X : \Omega \longrightarrow M_X$$
$$\omega \longmapsto X(\omega)$$

la variable qu'il faut estimer. On dispose pour cela de l'ensemble de variables $S = \{X_1, X_2, \ldots, X_r\}$, ou tout simplement, du vecteur S:

$$S : \Omega \longrightarrow M_S = M_1 \times M_2 \times \ldots \times M_r$$
$$\omega \longmapsto S(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_r(\omega))$$

11.2.1. - Cas non numérique

On peut imaginer deux critères pour qualifier une "bonne estimation":

a) Minimiser la probabilité d'erreur

On veut donc trouver la fonction

$$\hat{X}: M_S \longrightarrow M_X$$
$$s \longmapsto \hat{X}(s)$$

telle que

$$Pr(\hat{X}(S) \neq X) \leq Pr(f(S) \neq X), \quad , \forall f: M_S \longrightarrow M_X$$

Ce problème est résolu, et il est connu sous le nom d'*Observateur Idéal* [ASH 65, page 60] :

$$\hat{X}: M_S \longrightarrow M_X$$
$$s \longmapsto \hat{X}(s) = x_o$$
$$:\Leftrightarrow \quad Pr(X=x_o/S=s) \geq Pr(X=x/S=s) \quad \forall x \in M_X$$

autrement dit, la fonction $\hat{X}$ fait correspondre à chaque $s \in M_S$ la modalité x_o de X pour laquelle $P_r(X=x_o/S=s)$ est maximale.

b) Minimiser l'incertitude résiduelle

Dans ce cas, on veut trouver la fonction

$$\hat{X}: M_S \longrightarrow M_X$$
$$s \longmapsto \hat{X}(s)$$

telle que

$$H(X/\hat{X}(S)) \leq H(X/f(S)), \ \forall\ f:M_S \longrightarrow M_X$$

c'est-à-dire, telle que sachant $\hat{X}(S)$, ce qui nous reste à connaitre sur X soit minimal.

Nous avons essayé de caractériser de tels estimateurs dans le cas H = Entropie, sans avoir trouvé jusqu'à présent aucun résultat important. Pour des ensembles M_S et M_X de petite taille (Variables dichotomiques, par exemple), une recherche exhaustive de $\hat{X}$ serait envisageable.

11.2.2. - Variables Numériques

Si les variables X; X_1, X_2, ..., X_r prennent leurs valeurs dans l'ensemble des nombres réels $\mathbb{R}$, on tombe dans le domaine classique de l'Identification. Une famille importante de méthodes cherchent à déterminer la fonction :

$$\hat{X}: \mathbb{R}^r \longrightarrow \mathbb{R}$$
$$s \longmapsto \hat{X}(s)$$

appartenant à une classe $\mathcal{C}$ de fonctions de $\mathbb{R}^r$ vers $\mathbb{R}$, telle que

$$\left[\sum_{\omega\ \varepsilon\ \Omega} |X(\omega) - \hat{X}(S(\omega))|^{\alpha}\right]^{1/\alpha} \leq \left[\sum_{\omega\ \varepsilon\ \Omega} |X(\omega) - f(S(\omega))|^{\alpha}\right]^{1/\alpha}$$

et cela quelle que soit $f \varepsilon \mathcal{C}$. Lorsque $\alpha = 2$, ce problème est connu sous le nom de *Moindres Carrés* (Remarquons que dans le problème d'identification posé de cette façon, l'entropie (ou semi-valuation) H ne joue aucun rôle).

On peut aussi imaginer l'approche suivante :
on sait que l'entropie d'une variable constante est nulle et que intuitivement parlant, plus la variable change, plus grande sera son entropie. Alors, on pourrait chercher la variable $\hat{X} \varepsilon \mathcal{C}$ telle que

$$H(X - \hat{X}(S)) \leq H(X - f(S)) \quad \forall\ f \varepsilon \mathcal{C}$$

Ainsi si $H(X - \hat{X})$ est proche de zéro, on pourrait dire que $X - \hat{X}$ est "à peu près " constante, donc

$$X \approx \hat{X}(S) + K^{\underline{te}}$$

Nous n'avons pas trouvé de résultat significatif dans cette dernière voie jusqu'à présent.

Enfin, signalons pour terminer que R.E.RINK et J.DUFOUR [RINK 73; DUFO 79] ont aussi abordé le problème d'identification à l'aide de l'entropie, mais sous une optique un peu différente. Dans les idées que nous venons de présenter, l'entropie ou ses indices sont utilisés comme critère d'optimalité pour qualifier une "bonne estimation". Par contre, R.E.RINK et J.DUFOUR fixent à priori les types d'équations qui lient les variables, et expriment leurs différents coefficients en termes d'entropie et des indices qui en découlent.

12 - CONCLUSIONS

Dans ce chapitre nous avons essayé de formaliser des concepts tels que "Expliquer une variable à partir d'autres variables", "Partie Explicable", "Partie Modélisable", "Pourcentage encore non expliqué de la Partie Explicable", "Total non Explicable", ... Nous avons défini ce que sont les "Sous-ensembles Explicatifs Optimaux", et présenté quatre algorithmes pour les déterminer :

- AE : Algorithme Exhaustif,
  qui permet de trouver tous ces sous-ensembles optimaux;
- AA : Algorithme Agrégatif,
- AD : Algorithme Désagrégatif,
  qui permettent de trouver quelques uns des sous-ensembles optimaux, et d'approximer les autres; et enfin
- AAS : Algorithme d'Approximations Successives,
  qui trouve des sous-ensembles explicatifs vérifiant une condition nécessaire d'optimalité.

Pour avoir une idée du temps de calcul de ces algorithmes, nous avons étudié leur ordre. Ainsi, sous l'hypothèse d'un ordinateur qui réalise 1000 fois/seconde les instructions d'une boucle principale (un petit nombre d'opération +, -, et d'accès mémoire), et si l'on se permet au maximum 2 1/2 heures de temps de calcul, la taille maximale de Σ est donnée dans le tableau ci-dessous :

	AE	AA	AD	AAS
$\lvert\Sigma\rvert$=N Maximum	19	262	262	80

On peut dire que, exception faite de AE, les autres algorithmes peuvent être appliqués à des systèmes de taille assez élevée.

Nous avons exploré brièvement les applications possibles de cette approche ainsi que des différents concepts formalisés :

- utilisation pour qualifier le choix des variables de Σ, ainsi que pour évaluer les possibilités de modélisation mathématique.

- estimation de la faisabilité d'un problème de Reconnaissance de Formes, et détermination des paramètres les plus appropriés pour cette tâche de Reconnaissance.

- connaissant les Sous-ensembles Explicatifs Optimaux, construction d'un Graphe de Dépendance,qui permet d'étudier la structure des liaisons entre les variables. Ce graphe présente un caractère "multidimensionnel", qui prend en compte des aspects échappant aux graphes classiques construits sur des critères binaires. Bien évidemment, toutes les techniques disponibles d'analyse, hiérarchisation et visualisation sont applicables sur le Graphe de Dépendance.

- enfin, le dernier paragraphe a présenté le problème de l'Identification explicite des relations mathématiques liant les variables à l'aide de l'entropie. Nous avons signalé dans quels cas ce problème est déjà résolu, et proposé quelques voies de recherche possibles.

CONCLUSIONS GÉNÉRALES

Ce travail s'adresse principalement au praticien qui commence l'étude d'un système complexe. A ce stade, les seules connaissances dont il dispose en général, sont les variables qu'il croit pertinentes pour la description du système avec, peut être, quelques idées concernant leurs relations. La modélisation du système nécessite alors une phase d'Analyse Structurale à partir des données relevées sur celui-ci.

Tout d'abord, nous avons développé une forme de représentation qui s'adapte à tout type de système ,et qui présente l'avantage de rendre formellement analogues le cas statique et le cas dynamique. Cette représentation s'inscrit dans le cadre d'une Hiérarchie Epistémologique visant à organiser les différents niveaux de connaissances qu'un observateur peut avoir sur un système. Les diverses démarches de l'Analyse Structurale sont alors apparues clairement comme des déplacements dans cette hiérarchie.

Nous avons ensuite effectué une synthèse des principaux résultats de la Théorie de l'Information concernant les systèmes multi-variables. Deux aspects importants ont été mis en relief : d'une part, le très clair et fort appui intuitif de chacun des concepts et des résultats. D'autre part, ces résultats ont pu être classés selon quatre groupes d'hypothèses générales (Semi-Valuation, Semi-Valuation Monotone, Semi-Modulation Positive, Semi-Modulation Positive Monotone). L'Entropie n'étant pas la seule fonction à vérifier ces hypothèses, l'étiquette (SV, SVM, SMP, SMPM) qui accompagne chacun de nos résultats et algorithmes permet son application immédiate à n'importe quelle fonction vérifiant les hypothèses respectives. (On peut trouver plusieurs exemples de semi-valuations dans le chapitre I et IV de la thèse de Zohir ABID [ABID 79], ainsi qu'une intéressante méthode pour construire des semi-valuations à partir de n'importe quel Indice de Similarité (Annexe D). Reste à montrer lesquels des exemples donnés par Z.ABID vérifient aussi les hypothèses de Semi-Modulation). Enfin, il est important de remarquer que si la fonction utilisée est l'Entropie, aucune hypothèse algébrique n'est nécessaire ni sur les variables(qui peuvent alors être quantitatives, qualitatives, logiques ...) ni sur leurs relations (linéaires, non linéaires, ...).

Dans les deux derniers chapitres, nous avons étudié deux méthodes concrêtes d'Analyse Structurale :

- la Décomposition d'un système en sous-systèmes faiblement couplés : nous avons étudié en détail le concept de partition optimale pour proposer une approche qui réunit les cas statique et dynamique, et qui ne nécessite aucune des hypothèses de déterminisme requises par d'autres approches. Cinq algorithmes ont été étudiés, qui permettent de trouver ou d'approximer les décompositions optimales.

- la deuxième méthode est basée sur la détermination des sous-ensembles explicatifs optimaux. Cette méthode, qui avait été peu étudiée, possède d'intéressantes applications aux problèmes de Reconnaissance des Formes,d'Estimation de Variables et de détermination des interconnexions internes du système. Quatre algorithmes ont été proposés, qui permettent l'analyse de systèmes de taille assez élevée.

Loin d'être épuisé, le domaine que nous avons voulu explorer dans cette thèse demande encore bien des recherches. Signalons en particulier :

- la décomposition multicritère de systèmes : en effet, sur quelques exemples que nous avons traités, nous avons vu qu'il ne suffit pas toujours que les sous-systèmes soient faiblement couplés. Il peut être aussi nécessaire d'avoir des sous-systèmes d'une complexité uniforme, ou bien, que ces sous-systèmes respectent certaines limitations imposées (telles variables doivent forcément appartenir à un même sous-système, et telles autres à des sous-systèmes différents), ou encore, que la transinformation entre paires de sous-systèmes ne dépasse pas certaines limites, ... ;

- Développer des méthodes basés sur l'entropie visant à identifier explicitement les fonctions liant les variables. De telles méthodes auraient l'avantage d'être applicables à des variables qualitatives, cas pour lequel peu de résultats ont été obtenus.

- Il serait de même intéressant d'étudier la fiabilité et l'invariance des résultats que l'on obtient avec les différentes techniques de l'Analyse Structurale, par rapport à l'ensemble des données dont on dispose.

- Enfin, à plus long terme, le développement d'un outil informatique dans un langage de haut niveau, qui assure la mise en oeuvre des méthodologies étudiées.

Disons pour terminer que dans l'Analyse Structurale, il n'est pas question de proposer "la meilleure méthode", ou "la meilleure approche". Il s'agit seulement de fournir au praticien des outils généraux qui l'aident à simplifier et à orienter les premières phases de l'étude d'un système. Les méthodes proposées dans cette thèse demandent seulement que l'on dispose d'un bon ensemble d'observations sur le système,"... et l'on doit accumuler des observations partout où l'on veut faire oeuvre scientifique " [CAIL 76,pp.iv].

ANNEXES

ANNEXE 1

ORDRE DES ALGORITHMES DE DECOMPOSITION

	RE =	AH =	DR ≤	AS ≤	PD =
1	1	0	0	0	0
2	2	1	1	2	1
3	5	4	4	6	6
4	15	10	11	13	36
5	52	20	26	196	160
6	203	35	57	559	630
7	877	56	120	1236	2324
8	4140	84	247	2477	8232
9	21147	120	502	4764	28368
10	115975	165	1013	9043	95850
11	678570	220	2036	17188	319132
12	4'213597	286	4083	32929	1'050588
13	27'644437	364	8178	63708	3'427736
14	190'899322	455	16369	124391	11'102910
15	1'382'958550	560	32752	244692	35'749380
16	10'480'142200	680	65519	484021	114'529104
17	82'864'869800	816	131054	961180	365'340064
18	682'076'806000	969	262125	1'913755	1'161'081810
19	5'832'742'200000	1140	524268	3'816900	3'678'004300
20	51'724'158'200000	1330	1'048555	7'620905	11'617'372100

RE: Recherche Exhaustive
AH: Analyse Hiérarchique
DR: Division Recursive
AS: Approximations Succesives
PD: Programmation Dynamique.

ANNEXE 2

GÉNÉRALISATION DE L'ALGORITHME AH

L'algorithme ci-dessus utilise une pile, vide au départ, dans laquelle on va stocker des éléments de la forme [nombre entier; partition; nombre réel].

0. début;
1. Lire ε;
2. $M_N \leftarrow \{(X_1),(X_2),\ldots,(X_N)\}$;
3. $I_N \leftarrow \overleftrightarrow{I}(X_1:X_2:\ldots:X_N)$;
4. $r \leftarrow N$;
5. tant que $r \geqslant 3$ faire;
   - 5.1 Parmi tous les couples possibles de sous-systèmes de M_r trouver le couple R_α, R_β tel que $\overleftrightarrow{I}(R_\alpha:R_\beta)$ soit maximale; trouver aussi tous les couples R_u, R_v tels que $\left|\overleftrightarrow{I}(R_\alpha:R_\beta)-\overleftrightarrow{I}(R_u:R_v)\right| \leqslant \varepsilon$;
   - 5.2 Pour chacun des couples R_u, R_v trouvés, garder dans la pile $[r-1;(M_r \setminus \{R_u,R_v\})\cup\{R_u \cup R_v\}; I_r - \overleftrightarrow{I}(R_u:R_v)]$;
   - 5.3 $M_{r-1} \leftarrow (M_r \setminus \{R_\alpha,R_\beta\}) \cup \{R_\alpha \cup R_\beta\}$;
   - 5.4 $I_{r-1} \leftarrow I_r - \overleftrightarrow{I}(R_\alpha:R_\beta)$;
   - 5.5 $r \leftarrow r-1$;

   ftq;
6. $M_1 \leftarrow \{(X_1,X_2,\ldots,X_N)\}$;
7. $I_1 \leftarrow 0$;
8. Ecrire M_r, I_r, pour $r=1,2,\ldots,N$;
9. Si la pile n'est pas vide
   alors;
   - 9-1. $Z \leftarrow$ dernier élément de la pile;
   - 9-2. $r \leftarrow Z_1$ (première composante de Z);
   - 9-3. $M_r \leftarrow Z_2$ (deuxième composante de Z);
   - 9-4. $I_r \leftarrow Z_3$ (troisième composante de Z);
   - 9-5. aller au pas 5;

   fsi;
10. fin;

Cet algorithme écrit un certain nombre de chaines de partitions, le nombre de chaine étant dépendant de ε. Pour $\varepsilon=0$ il n'y aurait qu'une seule chaine (voire quelques ex-aequo). Pour un ε trop grand il y aurait trop de chaines (à la limite, toutes les chaines possibles menant de M_N à M_1, et il y en a énormément : $N!\ (N-1)!/2^{N-1}$).

Enfin, par un simple examen des différentes chaines écrites par l'algorithme on peut choisir pour chaque niveau la partition qui possède le plus petit $\overleftrightarrow{I}$.

ANNEXE 3

PRÉSENTATION DÉTAILLÉE DE L'ALGORITHME DR

Connaissant $M_r = \{R_1, R_2, \ldots, R_r\} \in \mathbb{P}_r(\Sigma)$, l'algorithme DR doit chercher le $R_\alpha \in M_r$ et le $S^* \subseteq R_\alpha$, $\emptyset \neq S^* \neq R_\alpha$, tels que

$$\overrightarrow{I}(R_\alpha \setminus S^* : S^*) \leq \overleftrightarrow{I}(R_i \setminus S : S) \quad , \forall R_i \in M_r, \forall\, S \subseteq R_i, \emptyset \neq S \neq R_i$$

Une fois trouvés R_α et S^*, on définit

$$M_{r+1} = (M_r \setminus \{R_\alpha\}) \cup \{R_\alpha \setminus S^*, S^*\} = \{R_1, R_2, \ldots, R_{\alpha-1}, R_\alpha \setminus S^*, S^*, R_{\alpha+1}, \ldots, R_r\}.$$

On voit alors que par rapport a M_r, il n'y a que deux classes qui ont changé dans M_{r+1}, ce sont $R_\alpha \setminus S^*$ et S^*. Toutes les **autres classes** de M_r ont déjà été examinées pour connaître leur meilleure partition en deux. Alors, pour ne pas tout recalculer, l'algorithme ci-dessous va stocker en permanence la meilleure partition en deux de chacune des classes de la partition en analyse.

Pour cela l'algorithme utilise trois vecteurs de travail, initialement vides, dont le contenu à l'étape r est :

CLASSE	PART	COUT
R_1	$R_1^{(1)}, R_1^{(2)}$	$\overleftrightarrow{I}(R_1^{(1)} : R_1^{(2)})$
R_2	$R_2^{(1)}, R_2^{(2)}$	$\overleftrightarrow{I}(R_2^{(1)} : R_2^{(2)})$
⋮	⋮	⋮
R_i	$R_i^{(1)}, R_i^{(2)}$	$\overleftrightarrow{I}(R_i^{(1)} : R_i^{(2)})$
⋮	⋮	⋮
R_r	$R_r^{(1)}, R_r^{(2)}$	$\overleftrightarrow{I}(R_r^{(1)} : R_r^{(2)})$

où $R_1,R_2,\ldots,R_r$ sont justement les classes de la partition $M_r = \{R_1,R_2,\ldots,R_r\}$. Pour chaque i, $PART_i$ contient les ensembles $R_i^{(1)},R_i^{(2)}$ qui sont la meilleure partition en deux de la classe R_i, c'est-à-dire

$R_i^{(1)}, R_i^{(2)} \subseteq R_i$; $R_i^{(1)} \cap R_i^{(2)} = \emptyset$; $R_i^{(1)} \cup R_i^{(2)} = R_i$; $R_i^{(1)} \neq \emptyset \neq R_i^{(2)}$ et $\overleftrightarrow{I}(R_i^{(1)}:R_i^{(2)}) \leqslant \overleftrightarrow{I}$ (n'importe quelle autre partition en deux de R_i).

Enfin, l'algorithme stocke les résultat dans les vecteurs M_r (pour les partitions) et I_r (pour l'intervaluation externe), i = 1,2,...,N.

```
. début
. M1 ← {(X1,X2,...,XN)};
. I1 ← 0;
. CLASSE1 ← (X1,X2,...,XN);
. r ← 1;
tant que r ⩽ N-2 faire;
    Pour i=1 à i=r faire;
        Si PARTi est vide
        alors;
            . Chercher la meilleure partition en deux de l'ensemble
              contenu dans CLASSEi, la placer dans PARTi, et dans COUTi
              mettre l'intervaluation respective (on va détailler cette
              instruction après); (*)
        fsi;
    fpour;
    . trouver le minimum entre COUT1, COUT2,...,COUTr. Soit α l'indice
      où se trouve ce minimum;
    . Mr+1 ← (Mr \ CLASSEα) ∪ PARTα
      (c'est-à-dire,construire Mr+1 à partir de Mr, en éliminant le
       sous-système contenu dans CLASSEα, et en mettant à sa place les
       deux sous-systèmes qui se trouvent dans PARTα. Bien entendu,
       Mr n'est pas changé);
```

```
    . I_{r+1} ← I_r + COUT_α;
    . j ← r;
    tant que j ⩾ α + 1 faire;                    ┐
       . CLASSE_{j+1} ← CLASSE_j;                │
       . PART_{j+1}   ← PART_j;                  │ décaler d'une position vers
       . COUT_{j+1}   ← COUT_j;                  │ "le bas" les vecteurs CLASSE,
       . j ← j-1;                                │ PART et COUT, à partir de
    ftq;                                         │ la position α + 1.
                                                 ┘
    . CLASSE_α ← Premier sous-ensemble de PART_α;
    . CLASSE_{α+1} ← Second sous-ensemble de PART_α;
    . Mettre "vide" dans les positions PART_α, PART_{α+1}, COUT_α, COUT_{α+1};
    . r ← r+1;
ftq;
. M_N ← {(X_1)(X_2)(X_3)...(X_N)};
. I_N ← I(X_1:X_2:...:X_N);
. Ecrire M_r, I_r, pour r=1,2,...,N;
. Fin;
```

Dans l'instruction (*) on cherche la meilleure partition en deux du sous-système contenu dans $CLASSE_i$. Si ce sous-ensemble n'a qu'un seul élément, une telle partition en deux est impossible et on mettra $COUT_i = +\infty$. Voici le morceau d'algorithme correspondant:

```
. COUT_i ← +∞;
Si | CLASSE_i | > 1
    alors;
        Pour tous les S ⊆ CLASSE_i, ∅ ≠ S ≠ CLASSE_i faire; (**)
            Si ↔I(CLASSE_i \ S:S) < COUT_i
                alors;
                    · COUT_i ← ↔I(CLASSE_i \ S:S);
                    · PART_i ← {CLASSE_i \ S,S};
            fsi;
        fpour;
fsi;
```

Remarque: en réalité dans l'instruction (**) ci dessus il n'y a pas besoin d'examiner tous les $S \subseteq CLASSE_i$, avec $\emptyset \neq S \neq CLASSE_i$. Il suffit d'en examiner la moitié puisque $\overleftrightarrow{I}$ est symétrique.

Enfin, l'algorithme DR présenté pourrait lui aussi être généralisé comme on l'a fait avec l'algorithme AH (Annexe 2), en lui faissant examiner, outre "la meilleure décomposition en deux", celles dont la différence avec la meilleure soit inférieure à un $\varepsilon > 0$ donné. Mais celà relève des details techniques de programmation.

ANNEXE 4

DÉMONSTRATION DE LA PROPOSITION IV.3 (PAGE 167)

Soient $M_X = \{a_1, a_2, \ldots, a_u\}$ et $M_S = \{b_1, b_2, \ldots, b_v\}$ les ensembles des modalités effectivement prises par les variables X et S, c'est-à-dire, telles que $Pr(X = a_i)$, $Pr(S = b_j) > 0$

① " $\Rightarrow$ "

Supposons donc que $H(X/S) = 0$. Ainsi, et d'après [II.18] on a :

$$0 = H(X/S) = \sum_{b_j \in M_S} Pr(S=b_j).\left[- \sum_{a_i \in M_X} Pr(X=a_i|S=b_j).\log Pr(X=a_i|S=b_j)\right].$$

Etant donné que $Pr(S=b_j) > 0$, on en déduit :

$$\forall j,\ 1 \leq j \leq v : - \sum_{a_i \in M_X} Pr(X=a_i|S=b_j).\ \log Pr(X=a_i|S=b_j)$$

L'expression $\alpha.\log \alpha$ ne s'annule que pour $\alpha=0$ (par définition) et pour $\alpha=1$, cette dernière égalité implique que, pour un j_o fixé, $P(X=a_i|S=b_{j_o})$ est toujours nul, sauf pour un certain i_o pour lequel $P(X=a_{i_o}|S=b_{j_o})$. Cet i_o est forcément unique puisque $\sum_i Pr(X=a_i|S=b_{j_o})=1$.

La fonction f est alors définie par :

$$f : M_S \longrightarrow M_X$$

$$b_{j_o} \longmapsto f(b_{j_o}) := a_{i_o} \Leftrightarrow Pr(X=a_{i_o}|S=b_{j_o})=1.$$

Montrons à présent que $X=f(S)$. Soit $\omega_o \in \Omega$. Posons $S(\omega_o) = b_{j_o}$. Alors $X(\omega_o) = a_{i_o}$, avec $P(X=a_{i_o}|S=b_{j_o})=1$. Finalement par définition de f, on a $f(b_{j_o}) = a_{i_o}$, donc

$$X(\omega_o) = a_{i_o} = f(b_{j_o}) = f(S(\omega_o)) = (foS)(\omega_o)$$

"$\Leftarrow$" . Supposons maintenant qu'il existe $f:M_S \longrightarrow M_X$ telle que $X=f(S)$. Soit $\omega_o \in \Omega$, et soient $X(\omega_o) = a_{i_o}$ et $S(\omega_o) = b_{j_o}$. Alors par hypothèse

$$a_{i_o} = X(\omega_o) = f(S(\omega_o)) = f(b_{j_o})$$

Ainsi,

$$Pr\,(X=a_i|S=b_{j_o}) = \begin{cases} 0 \text{ si } a_i \neq a_{i_o} = f(b_{j_o}) \\ 1 \text{ si } a_i = a_{i_o} = f(b_{j_o}), \end{cases}$$

donc

$$H(X|S) = \sum_{b_{j_o} \in M_S} Pr(S=b_{j_o}) \cdot \left[- \sum_{a_j \in M_X} Pr\,(X=a_j|Y=b_{j_o}) \cdot \log Pr\,(X=a_j|Y=b_{j_o}) \right] = 0$$

puisque chacun des termes de la somme interne est nul.

(F.D. ①)

②

$\Leftrightarrow$ $H(X|S) = H(X)$

$\Leftrightarrow$ $H(S,X) - H(S) = H(X)$

$\Leftrightarrow$ $0 = H(X) + H(S) - H(S,X) =: \overleftrightarrow{I}(S:X)$

donc, d'après la proposition II.8,

$\Leftrightarrow$ S et X sont statistiquement indépendants.

(F.D. ②)

ANNEXE 5

DÉMONSTRATION DE LA PROPOSITION IV.4 (SMPM) (PAGE 171)

Montrons que $H(X_1^E, X_2^E, \ldots, X_e^E) - \sum_{i=1}^{e} H(X_i^E | \mathcal{O}_i^E) \geqslant 0$

(pour alléger la notation, nous omettons à chaque fois le symbole E).

$$H(X_1,\ldots,X_e) - \sum_{i=1}^{e} H(X_i | \Sigma \setminus X_i)$$

$$= H(X_1,\ldots,X_e) - \left[\sum_{i=1}^{e} H(\Sigma) - \sum_{i=1}^{e} H(\Sigma \setminus X_i) \right]$$

$$= \sum_{i=1}^{e} H(\Sigma \setminus X_i) - \sum_{i=1}^{e} H(\Sigma) + H(X_1,\ldots,X_e) + \left(\sum_{i=1}^{e} H(X_i) - \sum_{i=1}^{e} H(X_i) \right)$$

$$= \sum_{i=1}^{e} \left[H(\Sigma \setminus X_i) + H(X_i) - H(\Sigma) \right] - \left[\sum_{i=1}^{e} H(X_i) - H(X_1,\ldots,X_e) \right]$$

$$= \sum_{i=1}^{e} \overleftrightarrow{I}(X_i : \Sigma \setminus X_i) - \overleftrightarrow{I}(X_1 : X_2 : \ldots : X_e)$$

en utilisant la proposition II.12 ② (SV), on obtient

$$= \overleftrightarrow{I}(X_1 : \Sigma \setminus X_1) + \overleftrightarrow{I}(X_2 : \Sigma \setminus X_2) + \ldots + \overleftrightarrow{I}(X_{e-1} : \Sigma \setminus X_{e-1}) + \overleftrightarrow{I}(X_e : \Sigma \setminus X_e)$$

$$- \overleftrightarrow{I}(X_1 : X_2, X_3, \ldots, X_e) - \overleftrightarrow{I}(X_2 : X_3, \ldots, X_e) - \ldots - \overleftrightarrow{I}(X_{e-1} : X_e)$$

$$= \sum_{i=1}^{e-1} \left(\overleftrightarrow{I}(X_i : \Sigma \setminus X_i) - \overleftrightarrow{I}(X_i : X_{i+1}, \ldots, X_e) \right) + \overleftrightarrow{I}(X_e : \Sigma \setminus X_e)$$

$$\geqslant 0$$

puisque chacun des termes de la somme est positif, étant donné que $\Sigma \setminus X_i \supseteq \{X_{i+1},\ldots,X_e\}$ (c.f. Proposition II.24(SMPM)).

(F.D.)

ANNEXE 6

ORDRE DES ALGORITHMES DE RECHERCHE DES SOUS-ENSEMBLES EXPLICATIFS OPTIMAUX (Cas Statique)

	AE	AA	AD	AAS
1	0	0	0	0
2	2	2	2	72
3	9	9	9	324
4	28	24	24	864
5	75	50	50	1.800
6	186	90	90	3.240
7	441	147	147	5.292
8	1.016	224	224	8.064
9	2.295	324	324	11.664
10	5.110	450	450	16.200
11	11.253	605	605	21.780
12	24.564	792	792	28.512
13	53.235	1.014	1.014	36.504
14	114.674	1.274	1.274	45.864
15	245.745	1.575	1.575	56.700
16	524.272	1.920	1.920	69.120
17	1'114.095	2.312	2.312	83.232
18	2'359.278	2.754	2.754	99.144
19	4'980.717	3.249	3.249	116.964
20	10'485.740	3.800	3.800	136.800

AE: Algorithme Exhaustif
AA: Algorithme Aggregatif
AD: Algorithme Dessagregatif
AAS: Algorithme d'Approximations Successives.

BIBLIOGRAPHIE

[ABID 79] : Z. ABID
" Contribution à l'Analyse Structurale par la théorie des Demi-Treillis"
Thèse de Troisième Cycle. Université Claude Bernard, Lyon 1
N° d'Ordre 892, Octobre 1979.

[ASH 65] : R. ASH
" Information Theory "
John WILEY & SONS, 1965.

[ASHB 56] : W.R. ASHBY
" Introduction to Cybernetics "
CHAMPAN and Hall.London.1956.

[ASHB 58] : W.R. ASHBY
" Requisite Variety and its applications to the Control of Complex Systems "
Cybernetica (Namur-Belgique), Vol.1, n°2, pp.83-99. 1958.

[ASHB 65] : W.R. ASHBY
" Measuring the Internal Informational Exchange in a System "
Cybernetica (Namur-Belgique), Vol.8, n°1, pp.5-22. 1965.

[ASHB 69] : W.R. ASHBY
" Two tables of Identities governing Information flows within large Systems "
Communication of the American Society of Cybernetics
Vol.1, n°2, pp.2-8. 1969.

[ASHB 70a]: W.R. ASHBY
" Information flows within Coordinated Systems "
Progress on Cybernetics : Proceedings of the 1st International Congres pp.57-64
J.Rose Editor - GORDON & BREACH, London 1970.

[ASHB 70b]: W.R. ASHBY; M.R. GARDNER
" Connectance of Large Dynamic (Cybernetic) Systems : Critical Values for Stability "
Nature, vol.228, pp.784. November 1970.

[ASHB 73] : W.R. ASHBY
" Some Peculiarities of Complex Systems "
Cybernetic Medecine, Vol.9, n°2, pp.1-8. 1973.

[BART 66] : R.G. BARTLE
" The Elements of Integration "
John WILEY & SONS. 1964.

[BIRK 64] : C. BIRKHOFF
" Lattice Theory "
American Mathematical Society. 4th edition.1964.

[BLAC 68] : N.M. BLACHMAN
" The Amount of Information that Y gives about X "
I.E.E.E. Trans. on Information theory. Vol.IT-14, n°1, pp.27-31.
Janvier 1968.

[BONA 81] : I.S. BONATTI; J. BERNUSSOU
" Une structure à deux niveaux pour la gestion temps réel de réseaux téléphoniques "
Proc. Congrès AFCET Automatique 1981. pp.93-105.
Nantes, 27-29 Octobre 1981.

[CAIL 76] : F. CAILLEZ; J.P PAGES
" Introduction à l'Analyse des Données "
Société de Mathématiques Appliquées et de Sciences Humaines
SMASH, 9, rue Duban; Paris, 1976.

[CALV 81] : J.L. CALVET
" Décomposition sur la base de la structure série-parallèle de systèmes Dynamiques "
Proc. Congrès AFCET Automatique 1981. pp.119-130
Nantes, 27-29 Octobre 1981.

[CARP 80a] : M.J. CARPANO; M. DELARCHE
" *Apport des techniques Graphiques Intéractives à l'Analyse Structurale de Systèmes* "
I : Méthodologie . II: Exemples de Réalisation et d'Application
R.A.I.R.O. : Vol.14 n°2, pp.205-222;
Automatique Vol.14,n°3, pp.271-292. 1980.

[CARP 80b] : M.J. CARPANO
" *Automatic Display for Hierarchised Graph for Computer-Aided Decision Analysis* "
I.E.E.E. Trans.on Systems, Man and Cybernetics
Vol. SMC-10, n°11, pp.705-715, Novembre 1980.

[CHOW 68] : C.K. CHOW; C.N. LIU
" *Approximating Discrete probability distribution with dependence trees* "
I.E.E.E. Trans. on Information theory.Vol.IT-14, n°3, pp.462-467.
May 1968.

[CNRS 77] : " *Théorie de l'Information : Développements récents et applications* "
Colloques Internationaux du Centre National de la Recherche Scientifique. N°276
Cachan, 4 à 8 Juillet 1977.

[CONA 69] : R.C. CONANT
" *The Information transfer required in Regulatory Processes* "
I.E.E.E. Trans. on Systems, Science and Cybernetics
Vol.SSC-5, n°4, pp.334-338. Octobre 1969.

[CONA 70] : R.C. CONANT; W.R. ASHBY
" *Every Good Regulator of a System must be a Model of that System*"
International Journal of Systems Science.Vol.1, n°2,pp.89-97,
1970.

[CONA 72] : R.C. CONANT
" Detecting Subsystems of a Complex System "
I.E.E.E. Trans. on Systems, Man and Cybernetics . Vol SMC-2, n°4, pp. 550-553, September 1972.

[CONA 73] : R.C. CONANT; J.B. STEINBERG
" Information Exchanged in grasshopper interactions "
Information and Control. Vol.23, pp.221-233. 1973.

[CONA 74] : R.C. CONANT
" Information flows in Hierarchical Systems "
International Journal of General Systems, Vol.1, n°1, pp.9-18. 1974.

[CONA 76] : R.C. CONANT
" Laws of Information wich Govern Systems "
I.E.E.E. Trans. on Systems, Man and Cybernetics
Vol.SMC-6, n°4, pp.240-255. Avril 1976.

[CONA 79] : R.C. CONANT
" Communication without a channel "
International Journal of General Systems, Vol.5, pp 93-98. 1979.

[CONA 80] : R.C. CONANT
" Structural Modelling using a simple Information Measure "
International Journal of Systems Science, Vol.11, n°6, pp.721-730. 1980.

[CRON 63] : J.N. CRONHOLM
" A general method of obtaining exact sampling probabilities of the Shannon - Wiener measure of information H "
Psychometrika, Vol.28, n°4, pp.405-413. 1963.

[DELA 79] : M. DELARCHE
" Quelques outils Infographiques pour l'Analyse Structurale de Systèmes "
Thèse Docteur-Ingénieur, U.S.M. Grenoble. Juin 1979.

[DIDA 79] : E. DIDAY et Collaborateurs
" Optimisation en classification Automatique "
Tome 1 et 2
Publication de l'INRIA, Octobre 1979.

[DUBU 80] : B. DUBUISSON; P. LAVISON
" Surveillance of a Nuclear Reactor by use of a Pattern Recognition Methodology "
I.E.E.E. Trans. on Systems, Man and Cybernetics
Vol. SMC-10, n°10, pp.603-609. October 1980.

[DUFO 76] : J. DUFOUR, G. GILLES, C.FOULARD
" Analyse Structurale et partition des Systèmes Dynamiques Complexes à l'aide de la Théorie de l'Information. Etude de l'influence du nombre de classes "
Comptes Rendus de l'Académie de Sciences de Paris
t.282, Série A, pp.543-546. 8 Mars 1976.

[DUFO 79] : J. DUFOUR
" Méthodes et Méthodologies d'Analyse de Systèmes Complexes. Application aux procédés Industriels et aux Systèmes Macroéconomiques "
Thèse d'Etat. Université Claude Bernard, Lyon. n° D'Ordre 79.17
Mai 1979.

[FIND 80] : W. FINDEISEN, et al.
" Control and Coordination in Hiérarchical Systems "
John Wiley and Sons. 1980.

[FOSS 81] : A.J. FOSSARD; J.F. MAGNI
" Modélisation, Commande et applications des Systèmes à échelles de temps multiples "
Proc. Congrès AFCET Automatique 1981. pp.33-55
Nantes, 27-29 Octobre 1981.

[GABA 78] : D. GABAY; P. NEPOMIASTCHY, M. RACHDI, A. RAVELLI
" Etude, Résolution et Optimisation de Modèles Macroéconomiques "
Rapport de Recherche R-312
Laboria INRIA. Juin 1978.

[GARN 56] : W.R. GARNER, N.J. MCGILL
" The Relation between information and variance Analysis"
Psychometrica, Vol.21, pp.219-228 . 1956.

[GUIA 77] : S. GUIASU
" Information Theory with Applications "
Mc-Graw Hill, London, 1977.

[KICK 78] : W.J.M. KICKERT, J.W.M. BERTRAN, J. PRAAGMAN
" Some Comments on Cybernetic and Control "
I.E.E.E. Trans. on Systems, Man and Cybernetics,
Vol. SMC-8, n°11, pp.805-809. Novembre 1978.

[KLIR 69] : G.J. KLIR
" An Approach to General Systems Theory "
Van Nostrand Reinhold, New-York, 1969.

[KLIR 75] : G.J. KLIR
" On the Representation of Activity Arrays "
International Journal of General Systems, Vol.2, n°3,
pp.149-168, 1975.

[KLIR 76] : G.J. KLIR
" Identification of Generative Structures in Empirical Data "
International Journal of General Systems, Vol.3,
pp.89-104, 1976.

[KLIR 77] : G.J. KLIR
" On the problem of computer-aided structure identification : some experimental observations and resulting guidelines "
International Journal of Man-Machine Studies
Vol.9, pp.593-628. 1977.

[KOUV 76] : D.D. KOUVATSOS
" Decomposition Criteria for the Design of Complex Systems "
International Journal of Systems Sciences, Vol.7, n°10,
pp.1081-1088. 1976.

[KU 69] : H.H. KU, S. KULBACK
" Approximating Discrete probability distribution "
I.E.E.E. Trans. on Information Theory.Vol.IT-15, pp.444-447,
July 1969.

[LANG 81] : B. LANG, N.E. CHEICKH OBEID
" Régulation optimale de systèmes complexes continus utilisant une méthode de décomposition-coordination par relaxation "
Proc.Congrès AFCET Automatique 1981. pp.107-117
Nantes, 27-29 Octobre 1981.

[LARM 75] : P. LARMINANT, Y. THOMAS
" Automatique des Systèmes Linéaires "
Tome 1 : Signaux et Systèmes
Flammarion Sciences. 1975.

[LERM 76] : I.C. LERMAN
" Reconnaissance et Classification des Structures finies en Analyse de Données "
Vol.1. Théorie et Méthodes. Rapport n°70.
IRISA, Institut de Recherche en Informatique et Systèmes Aléatoires
B.P. 25A; 35031 Rennes Cédex.1976.

[LERM 81] : I.C. LERMAN
" Classification et Analyse Ordinale des Données "
Dunod, Paris, 1981.

[LOPE 77] : R. LOPEZ DE MANTARAS BADIA
" Auto apprentisage d'une partition. Application au classement Itératif de Données Multidimensionnelles "
Thèse Doctorat de Spécialité. Université Paul Sabatier. Toulouse
N° d'Ordre 1998. Juin 1977.

[LOSF 74] : J. LOSFELD
" Information fournie par un enxemble d'informateurs et applications aux questionnaires et à l'Analyse de Données "
Thèse d'Etat en Mathématiques
Université de Lille 1, n°d'Ordre 304. Juin 1974.

[MAGN 81] : J.F. MAGNI, A.J. FOSSARD
" Commande en deux étapes des systèmes linéaires à deux dynamiques "
Proc. Congrès AFCET Automatique 1981. pp.131-141
Nantes, 27-29 Octobre 1981.

[MCGI 54] : W.J. MCGILL
" Multivariate Information Transmission "
Psychometrika, Vol.19, pp.97-116. 1954.

[MILL 55] : G.A. MILLER
" On the bias of Information Estimates "
in "Information Theory in Psychology", edited by H.Quastler.
pp.95-100. the Free Press.Glencoe, Illinois.1955.

[MILL 63] : G.A. MILLER
" What is Information Measurement "
American Psycologist, Vol.8, n°2, pp.50-51, 1963.

[MORG 81] : S.D. MORGERA
" Structured Estimation. Part II : Multivariate Probability Density Estimation "
I.E.E.E. Trans. on Information Theory.
Vol.IT-27, N°5, pp.607-622. September 1981.

[NEPO 78] : P.NEPOMIASTCHY,A. REVELLI, F. RECHENMAN
" An Automatic Method To get an econometric model in quasi triangular form "
Rapport de Recherche R-313
Laboria, INRIA, Juin 1978.

[PARZ 60] : E. PARZEN
" Modern Probability theory and its applications "
John Wiley and Sons. New-York.1960.

[PFAF 71] : E. PFAFFELHUBER
" Error Estimation for the determination of entropy and Information rate from relative frecuencies "
Kybernetics, Vol.8, pp.50-51, 1971.

[PORT 76] : B. PORTER
" Requisite variety in the Systems and Control Sciences "
International Journal of General Systems, Vol.2 pp 225-229.1976.

[RICH 75] : M. RICHETIN
" Analyse Structurale des Systèmes Complexes en vue d'une commande Hiérarchisée "
Thèse d'Etat. Université Paul Sabatier. Toulouse
N° d'Ordre 674. Juillet 1975.

[RINK 73] : R.E. RINK
" Information Theoretic Methods for Modelling and Analysing Large Systems "
5th IFIP Conference on Optimisation Techniques.Rome 1973.
Publié par Lecture Notes on Computer Science.N°3.
Springer Verlag. Allemagne.1973.

[RINK 75] : R.E. RINK, A.N. HALTER
" Mutual Entropy Interaction Between Growth Patterns of Oregon Counties "
I.E.E.E. Trans. on Systems, Man and Cybernetics
Vol.SMC-5, N°5, pp.399-402. Mai 1975.

[SALL 79] : J. SALLATIN
" Représentations d'observations dans le contexte de la théorie de l'Information "
Thèse d'Etat. Université Pierre et Marie Curie
Paris, 14 Juin 1979.

[SHAN 48] : C.E. SHANNON
" A Mathematical Theory of Communication "
Bell System Technical Journal, Vol.27, pp.379-423 et 623-656.1948.
Reimpressions :
- C.E. SHANNON, W.WEAVER
" The Mathematical theory of Communication "
the University of Illinois Press,Urbana, 1949
- aussi dans [SLEP 74]

[SIMO 62] : H.A. SIMON
" The Architecture of Complexity "
Proc. of the American Philosophical Society, Vol.106, n°6 décembre 1962.
réimprimé : General Systems, Vol.10, pp 63-76. 1965.

[SING 78] : M.G. SINGH, A. TITLI
" Systems : Décomposition, Optimisation and Control "
Pergamon Press. 1978.

[SLEP 74] : D. SLEPIAN, Editeur
" Key papers in the development of Information Theory "
I.E.E.E. Press, New-York. Réf PC00299. 1974.

[STAR 79] : M. STAROSWIECKI
" Contribution à l'Analyse et à la Commande de Systèmes Complexes à Critères Multiples "
Thèse d'Etat. Université Lille 1.
N° d'Ordre 444. Janvier 1979.

[STAR 81] : M. STAROSWIECKI, V.TORO
" Structural Analysis of Complex Systems by Means of Canonical Analysis "
VII International Conference on Systems Science
Wroclaw, Pologne, 15-18 Septembre 1981.

[STEI 74] : J. STEINBERG , R.C. CONANT
" An informational analysis of the intermale behavoir of the grasshopper "
Animal Behavoir, Vol.22. 1974.

[STEW 65] : D.V. STEWARD
" Partitioning an Tearing Systems of Equations "
Journal SIAM of Numerical Analysis. Série B
Vol.2, n°2, 1965.

[TITL 79] : A. TITLI et al.
" Analyse et Commande des Systèmes Complexes "
Monographie AFCET, Cepadues Editions. 1979.

[TORO 78] : V.TORO
" Identification de Subsistemas de Sistemas Complejos con base en la Teoria de la Informacion "
Tesis ISC-77-II-18. Departamento de Sistemas
Universidad de Los Andes, Bogotá (Colombie).1978.

[TORO 79] : V.TORO
" Décomposition des Systèmes Complexes "
Mémoire de DEA. Laboratoire d'Automatique
Université de Lille 1. Juin 1979.

[TORO 81] : V.TORO, M. STAROSWIECKI
" Méthodes Heuristiques et Décomposition Optimale des Systèmes Complexes "
Congrès AFCET Automatique 1981
Nantes, 27-29 Octobre 1981.

[WAGN 75] : T.J. WAGNER
" Non Parametric Estimates of Probability Densities "
I.E.E.E. Trans. on Information Theory, Vol.IT-21, n°4,
pp.438-440, 1975.

[WARF 73a] : J.W. WARFIELD
" On arranging elements of a Hierarchy in graphic form"
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol.SMC-3, N°2, pp.121-132 March 1973.

[WARF 73b] : J.W. WARFIELD
" Binary Matrices on System Modeling "
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol.SMC-3, N°2, pp.441-449. September 1973.

[WARF 74a] : J.W. WARFIELD
" *Developping Subsystems Matrices in Structural Modeling*"
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol.SMC-4, N°1, pp.74-80 January 1974.

[WARF 74b] : J.W. WARFIELD
" *Developing Interconnexion Matrices in Structural Modeling* "
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol.SMC-4, N°1, pp.81-87 January 1974.

[WARF 74c] : J.W. WARFIELD
" *Toward Interpretation of Complex Structural Models* "
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol. SMC-4, N°5, pp.405-417, September 1974.

[WARF 76] : J.W. WARFIELD
" *Implication Structures for Systems Interconnexion Matrices* "
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol.SMC-6, N°1, pp.18-24. January 1976.

[WARF 77] : J.W. WARFIELD
" *Crossing Theory and Hierarchy Mapping* "
I.E.E.E. Trans. on Syst. Man and Cybern.
Vol.SMC-7, N°7, pp.505-523 July 1977.

[WEID 69] : H.L. WEIDEMANN
" *Entropy Analysis of Feedback Control Systems* "
Advances in Control Systems. Theory and Application, pp.225-255
Vol.7 Edited by C.T. LEONDES
Academic Press, New-York.1969.

[WYNE 81] : A.D. WYNER
" *Fundamental limits in Information Theory* "
Proceedings of the I.E.E.E. Vol.69, n°2, pp.239-251. Fèvrier 1981

[YOUN 78] : T.Y. YOUNG, P.S. LIU
" *Linear transformation of binary random vector and its application to approximating probability distribution* "
I.E.E.E. Trans. on Information Theory.Vol.IT-24, n°2, pp.152-156, Mars 1978.

[YVIN 78] : B.YVINEC, M. MILGRAM, B. DUBUISSON
" *Estimation de l'Entropie dans le cas continu en vue de la décomposition des systèmes complexes* "
R.A.I.R.O. Automatique, Vol.12, n°1, pp.23-42, 1978.