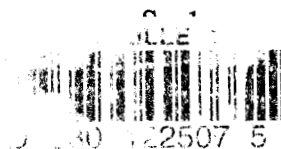


50376
1987
131

50376
1987
131

SOMMAIRE

INTRODUCTION GENERALE	i - 1
CHAPITRE I : ANALYSE DES SYSTEMES	
I - INTRODUCTION	I - 2
II - DESCRIPTION D'UN SYSTEME COMPLEXE	I - 3
<u>II.1 Description</u>	I - 3
<u>II.2 Variables et partitions</u>	I - 5
II.2.1 Le inf-demi treillis des partitions de Ω	I - 5
II.2.2 Le sup-demi treillis des variables vectorielles	I - 8
<u>II.3 Estimation des probabilités</u>	I - 9
<u>II.4 Recodage des variables</u>	I - 10
II.4.1 Position du problème	I - 11
<u>II.4.1.1 Cas des variables ordinales</u>	I - 11
<u>II.4.1.2 Cas des variables quantitatives</u>	I - 12
II.4.2 Caractérisation de $P_j^*(x)$:	I - 12
<u>II.4.2.1 Critères relatifs aux caractéristiques des classes :</u>	I - 13
<u>II.4.2.2 Critères dépendant du type de variables</u>	I - 14
III - OUTILS UTILISES POUR L'ANALYSE D'UN SYSTEME COMPLEXE	I - 14
<u>III.1 Les outils géométriques de l'analyse de données</u>	I - 14
III.1.1 Analyse en composantes principales	I - 15
III.1.2 Analyse factorielle des correspondances	I - 16
III.1.3 Régression linéaire	I - 16
III.1.4 Analyse canonique	I - 17



<u>III.2 Les outils de la théorie de l'information</u>	I - 18
III.2.1 Introduction	I - 18
III.2.2 Mesure d'information et entropie	I - 19
III.2.3 Une autre approche pour la définition de l'entropie de SHANNON d'une variable discrète :	I - 20
III.2.4 Les propriétés de l'entropie de SHANNON	I - 22
III.2.5 Les propriétés de l'entropie de Max :	I - 23
III.2.6 Entropie d'une variable continue	I - 24
III.2.7 Pourquoi s'intéresse-t-on à l'entropie de SHANNON ?	I - 27
IV - APPLICATION DES OUTILS DE L'ANALYSE DES DONNES ET DE LA THEORIE DE L'INFORMATION A L'AUTOMATIQUE	I - 29
<u>IV.1 L'apport des outils de l'analyse des données</u>	I - 29
IV.1.1 L'apport des méthodes factorielles	I - 29
IV.1.2 L'apport des méthodes de régression :	I - 30
IV.1.3 L'apport de l'analyse canonique	I - 31
<u>IV.2 L'apport des outils de la théorie de l'information</u>	I - 31
IV.2.1 Les transinformations	I - 31
IV.2.2 Les propriétés des transinformations	I - 32
IV.2.3 Entropie et transinformation conditionnelle	I - 35
<u>IV.2.3.1 Définitions</u>	I - 35
<u>IV.2.3.2 Quelques propriétés de l'entropie conditionnelle et des transinformations</u>	I - 37
IV.2.4 Le principe de conditionnement uniforme [GARNER 56, ASHBY 65]	I - 38
IV.2.5 Interprétation graphique d'un système :	I - 39
IV.2.6 Analyse structurale et indices de couplage	I - 43
<u>IV.2.6.1 Pourquoi s'intéresse-t-on à la T.I. pour l'analyse des systèmes ?</u>	I - 44
<u>IV.2.6.2 Indices de couplage</u>	I - 45
IV.2.7 Problèmes posés par les indices de couplage en présence du bruit : approche globale et approche locale	I - 51

<u>IV.2.7.1 Position du problème</u>	I - 51
<u>IV.2.7.2 Formulation du problème</u>	I - 52
<u>IV.2.7.3 Approche globale, approche locale</u>	I - 54
V - CONCLUSION	I - 55

CHAPITRE II : MODELISATION STRUCTURALE DES SYSTEMES COMPLEXES
RECHERCHE DE RELATIONS FONCTIONNELLES

I - INTRODUCTION	II - 2
II - ANALYSE INFORMATIONNELLE ET MODELISATION	II - 3
<u>II.1 Position du problème</u>	II - 3
<u>II.2 Insuffisance des approches classiques</u>	II - 4
II.2.1 Insuffisance de l'analyse binaire	II - 4
II.2.2 Insuffisance de l'analyse multidimensionnelle	II - 5
<u>II.3 Concept de modélisabilité</u>	II - 5
III - MODELISATION DANS LE CAS D'UN SYSTEME DETERMINISTE :	
DETERMINATION DES SOUS ENSEMBLES EXPLICATIFS	II - 6
<u>III.1 Formulation du problème</u>	II - 6
<u>III.2 Quelques résultats</u>	II - 7
<u>III.3 Mise en œuvre de l'algorithme de recherche des sous-ensembles explicatifs</u>	II - 9
<u>III.4 Les degrés de liberté du système et les sous-ensembles explicatifs multiples d'une variable</u>	II - 11
III.4.1 Degrés de liberté d'un système	II - 11
III.4.2 Les sous-ensembles explicatifs multiples	II - 14
III.4.3 Remarques	II - 16
IV - MODELISATION DANS LE CAS GENERAL	II - 16

<u>IV.1 Position du problème</u>	II - 16
<u>IV.2 Approche préliminaire</u>	II - 17
IV.2.1 - $H(Y/X) = H(Y)$:	II - 17
IV.2.2 - $H(Y/X) = 0$:	II - 18
IV.2.3 - $0 < H(Y/X) < H(Y)$	II - 18
<u>IV.3 Les approches classiques</u>	II - 19
IV.3.1 Les approches générales	II - 20
IV.3.2 Les approches utilisant un modèle	II - 20
<u>IV.4 L'approche proposée</u>	II - 22
IV.4.1 Insuffisance des approches classiques	II - 22
IV.4.2 Test de détection des modalités mal expliquées	II - 23
<u>IV.4.2.1 Les tests statistiques</u>	II - 23
<u>IV.4.2.2 Variation entropique</u>	II - 25
IV.4.3 Recherche des variables explicatives supplémentaires	II - 26
IV.4.4 Algorithmes	II - 27
<u>IV.4.4.1 Algorithme 2 : cas des variables secondaires disponibles</u>	II - 27
<u>IV.4.4.2 Synthèse structurale</u>	II - 29
 V - APPLICATION : MODELISATION D'UNE STATION D'EPURATION DES EAUX USEES UTILISANT LA VOIE BIOLOGIQUE	 II - 32
<u>V.1 Description d'une station d'épuration</u>	II - 32
<u>V.2 Principe de fonctionnement du système</u>	II - 33
<u>V.3 Les grandeurs d'entrée et de sortie</u>	II - 34
<u>V.4 Pilote et bilan des grandeurs accessibles</u>	II - 36
<u>V.5 Modélisation structurale : approche globale et approche locale</u>	II - 39
V.5.1 Constitution de la population d'apprentissage	II - 39
V.5.2 Analyse globale et analyse locale	II - 40
<u>V.5.2.1 Analyse globale</u>	II - 40
<u>V.5.2.2 Analyse locale</u>	II - 41
 VI - CONCLUSION	 II - 43

CHAPITRE III : MODELISATION DES SYSTEMES COMPLEXES
 APPROCHE SYSTEME EXPERT

I - INTRODUCTION	III - 2
II - NOTATIONS COMPLEMENTAIRES	III - 3
III - QUELLE CONNAISSANCE PEUT-ON EXTRAIRE DE [Pij]	III - 4
<u>III.1 Définition</u>	III - 5
<u>III.2 Problèmes posés lors de la formulation des connaissances</u>	III - 5
<u>III.3 Nature des règles</u>	III - 6
III.3.1 Les règles de type 1	III - 6
III.3.2 Les règles de type 2	III - 7
IV - EXTRACTION DE CONNAISSANCES : APPROCHE UTILISANT UN OPERATEUR DE CONTRASTE	III - 7
<u>IV.1 Opérateur de contraste</u>	III - 8
<u>IV.2 Analyse du niveau de contraste</u>	III - 12
<u>IV.3 Quelles connaissances peut-on extraire d'une ligne après l'opération de contraste ?</u>	III - 15
<u>IV.4 Critères d'évaluation des seuils de contraste</u>	III - 16
IV.4.1 Position du problème	III - 16
IV.4.2 Jugement du contraste en termes de probabilité	III - 19
IV.4.3 Jugement du contraste en termes de perte d'information	III - 20
<u>IV.4.3.1 Le premier modèle</u>	III - 20
<u>IV.4.3.2 Les autres modèles de comportement des résidus</u>	III - 22
A - Ensemble de modalités non structurées	III - 23
B - Ensemble de modalités structurées	III - 24
<u>IV.4.3.3 Relation fondamentale</u>	III - 25
<u>IV.4.3.4 Définition de $U(\sigma)$</u>	III - 26

IV.4.4 Jugement en termes d'efficacité	III - 29
<u>IV.5 Coefficients de vérité et de non vérité d'une règle</u>	III - 32
IV.5.1 Introduction	III - 32
IV.5.2 Coefficient de vérité d'une règle	III - 33
V - EXTRACTION DE CONNAISSANCES : APPROCHE ANALYSE STRUCTURALE PAR LA THEORIE DE L'INFORMATION	III - 35
<u>V.1 Introduction</u>	III - 35
<u>V.2 Répartition des résidus</u>	III - 36
<u>V.3 Exploitation de l'information contenue dans une ligne</u>	III - 39
<u>V.4 Coefficient d'extraction de l'information</u>	III - 42
V.4.1 Définitions	III - 42
V.4.2 Etude de E(s) :	III - 43
<u>V.5 Les critères de jugement du niveau d'extraction limite</u>	III - 45
V.5.1 Jugement en terme d'utilité de l'information non extraite	III - 45
V.5.2 Jugement en termes de probabilité d'occurrence :	III - 46
V.5.3 Jugement en termes de réduction d'information	III - 46
V.5.4 Jugement en termes d'efficacité	III - 46
<u>V.6 Coefficient de vérité</u>	III - 47
<u>V.7 Comparaison des deux approches</u>	III - 48
VI - COMPARAISON DES DEUX APPROCHES AVEC L'APPROCHE ANALYSE DES DONNEES	III - 49
<u>VI.1 Introduction</u>	III - 49
<u>VI.2 Analyse des tableaux contrastés</u>	III - 49
<u>VI.3 Analyse de $[\phi_{ij}]_P$</u>	III - 50
<u>VI.4 Récapitulation</u>	III - 51
VII - CONCLUSION	III - 51

CHAPITRE IV : SIMPLIFICATION DU MODELE OBTENU
PAR L'APPROCHE SYSTEME EXPERT

I - INTRODUCTION	IV - 2
II - REGROUPEMENT DES LIGNES DU TABLEAU INITIAL	IV - 3
<u>II.1 Perte d'information provoquée par l'opération de regroupement des lignes initiales</u>	IV - 3
II.1.1 Utilisation de la distance du χ^2	IV - 3
II.1.2 Utilisation de l'information de SHANNON	IV - 4
II.1.3 Comparaison de ΔI et $\Delta \chi^2$	IV - 5
<u>II.2 Coefficient de vérité de la règle composée</u>	IV - 6
<u>II.3 Coefficient d'efficacité de la règle composée</u>	IV - 6
III - OPERATION DE CONTRASTE APRES AGREGATION DES LIGNES	IV - 7
<u>III.1 La nouvelle distribution contrastée :</u>	IV - 7
<u>III.2 Coefficient de vérité de la règle composée</u>	IV - 8
<u>III.3 Coefficient d'efficacité de la règle composée</u>	IV - 9
IV - REGROUPEMENT DES LIGNES DU TABLEAU CONTRASTE	IV - 9
<u>IV.1 La nouvelle distribution contrastée</u>	IV - 9
<u>IV.2 Perte d'information occasionnée par le regroupement</u>	IV - 10
<u>IV.3 Coefficient de vérité de la règle composée</u>	IV - 11
<u>IV.4 Coefficient d'efficacité de la règle composée</u>	IV - 13
V - COMPARAISON DES DEUX PROCEDURES	IV - 15
<u>V.1 Comparaison de RG o Cσ et Cδ o RG</u>	IV - 15

VI - SIMPLIFICATION DES REGLES COMPOSEES	IV - 17
<u>VI.1 Introduction</u>	IV - 17
<u>VI.2 Les techniques habituelles de simplification</u>	IV - 18
VI.2.1 Les techniques graphiques [tableaux de KARNAUGH]	IV - 18
VI.2.2 Les techniques algébriques ou systématiques	
[méthode de QUINE MC CLUSKEY, principe PRESTO]	IV - 18
<u>VI.3 Méthode entropique de simplification</u>	IV - 19
VI.3.1 Exemple	IV - 20
VI.3.2 Méthode de simplification	IV - 21
VI.3.3 Exemple d'application	IV - 24
<u>VI.4 Algorithme de simplification du modèle</u>	IV - 26
<u>VI.5 Modélisation de Y : explication d'une colonne par l'ensemble</u>	
<u>des lignes contrastées</u>	IV - 29
VII - APPLICATION : AIDE A LA DECISION AU DIAGNOSTIC MEDICAL	IV - 32
<u>VII.1 Position du problème</u>	IV - 32
<u>VII.2 Mise en œuvre du processus d'aide à la décision</u>	IV - 33
VII.2.1 Introduction	IV - 33
VII.2.2 Les différents comportements du médecin	IV - 35
VII.2.3 Aspect conversationnel : analyse des différentes situations	IV - 36
VIII - CONCLUSION	IV - 37
CONCLUSION GENERALE	C - 1
ANNEXES	A - 1
BIBLIOGRAPHIE	B - 1

INTRODUCTION GENERALE

Les travaux présentés dans cette thèse se situent dans le cadre de la modélisation des systèmes complexes. Deux approches sont envisagées :

- . l'approche analyse structurale : recherche de relations fonctionnelles ;
- . l'approche système expert : extraction des connaissances pour la validation des règles d'un système expert.

Dans un premier chapitre, on montre que l'analyse des systèmes sur lesquels on ne possède aucune information à priori peut être abordée de deux façons différentes : d'une part, les méthodes géométriques de l'analyse des données et d'autre part, les concepts de l'analyse structurale issus de la théorie de l'information. Nous présentons plus particulièrement les résultats de cette dernière approche puisque celle-ci se prête bien à l'analyse des systèmes et possède l'avantage d'être sensible à tout type de relation. On montre que les indices de similarité classiques et les indices informationnels, conçus indépendamment les uns des autres, peuvent être harmonisés sous un même formalisme. Or en présence du bruit sur les données brutes recueillies sur le système, l'utilisation de ces indices, pour effectuer une approche globale de modélisation, peut s'avérer insuffisante et peut conduire à des erreurs d'analyse. Pour résoudre un tel problème, on propose l'utilisation des approches locales qui consistent à travailler sur des sous-échantillons de la population d'apprentissage de taille plus réduite.

Au chapitre II consacré à la modélisation structurale, on étudie dans un premier temps le concept de modélisabilité, on cherche pour chaque variable du système des sous-ensembles explicatifs optimaux. On montre ensuite certaines propriétés de ces sous-ensembles qui permettent d'accélérer notablement le processus de leur détermination. On présente à cet effet un algorithme dans le cas déterministe. On met en évidence les notions de sous-ensembles explicatifs multiples d'une variable et les degrés de liberté d'un système qui se traduisent par la recherche des relations redondantes.

Pour la modélisation du système dans le cas général, on montre qu'en l'absence d'une description complète de celui-ci et/ou en présence du bruit sur ses modalités, une analyse binaire ou multi-dimensionnelle utilisant les indices informationnels classiques est insuffisante pour son explication totale. Dans ce sens, on propose la mise en œuvre de tests statistiques et on développe un indice basé sur la variation entropique ce qui conduit à distinguer les modalités du système bien expliquées et les modalités du système dont l'explication requiert l'introduction de variables explicatives supplémentaires. Cette recherche peut s'effectuer parmi les variables secondaires (disponibles ou candidates) pour décrire le système mais peut aussi se traduire par la création de nouvelles variables (synthèse structurale). On présente à cet effet, un algorithme permet-

tant l'obtention de ces sous-ensembles explicatifs optimaux. On développe une application au domaine de l'épuration des eaux usées et on montre l'intérêt de l'analyse structurale utilisant des approches locales pour la modélisation d'un tel système.

Le troisième chapitre étudie une nouvelle méthode de modélisation des systèmes complexes visant l'extraction de connaissances sous forme de règles à partir d'un tableau de contingence. On développe à cet effet, deux approches similaires : la première est basée sur la définition d'un opérateur de contraste, la deuxième utilise une procédure pas à pas basée sur la définition d'un coefficient d'extraction de l'information. Dans les deux cas, le choix du niveau de contraste ou du seuil d'extraction à retenir est opéré en fonction d'une analyse multi-critère faisant intervenir les notions de probabilité d'occurrence, de perte d'information, de conformité des données à la réalité, d'efficacité et de vérité des règles à extraire. On montre que les deux approches aboutissent aux mêmes résultats. Les règles extraites respectivement des lignes et des colonnes du tableau de contingence sont dotées d'un coefficient de vérité et d'un coefficient d'efficacité et constituent ensemble le modèle redondant du système (règles atomiques). L'utilisation d'une telle approche conjointement avec la méthode d'analyse factorielle des correspondances peut faciliter la représentation et l'expression des connaissances à partir d'un tableau de contingence.

Au chapitre IV, on se propose de simplifier le modèle redondant du système obtenu par l'approche précédente. En effet, lorsque les règles aboutissent à des conclusions identiques ou voisines, il est possible de procéder à leur regroupement, c'est-à-dire d'agréger les lignes et les colonnes. Cette façon de procéder conduit lorsque cette opération est possible, à des règles composées pour lesquelles il convient d'étudier la perte d'information qu'elles occasionnent et le gain de simplicité qu'elles permettent. On montre globalement que la perte d'information occasionnée par l'opération de contraste se traduit par un gain d'efficacité alors que l'opération de regroupement provoque une perte de celle-ci. On présente à cet effet un algorithme qui permet d'agréger les règles dont l'efficacité ne se trouve pas trop dégradée et dont la partie hypothèse s'exprimera sous une forme plus simple. On montre que l'entropie du Max conduit dans certains cas à la simplification de ces règles. L'ensemble des règles obtenues, constitue le modèle simplifié du système qui peut être modifié ou mis à jour à chaque fois que de nouvelles données sont disponibles.

L'approche proposée constitue un extracteur de connaissance pouvant être utilisé en modélisation reconnaissance des formes, les divers domaines du diagnostic, la mise à jour et la vérification des règles d'un système expert.

On développe une application dans le domaine de l'aide à la décision médicale. L'objectif dans ce cas est de proposer au médecin, compte tenu des données disponibles, les conclusions tirées des règles extraites de la population d'apprentissage.

CHAPITRE I

ANALYSE DES SYSTEMES

I - INTRODUCTION

Les domaines d'application de "l'approche systèmes" ne cessent de s'élargir pour inclure aujourd'hui des systèmes technologiques, écologiques, biologiques, socio-économiques... Parallèlement, la complexité des systèmes pris en compte s'accroît et pose de sérieux problèmes d'analyse. Cette augmentation de la complexité, à laquelle s'ajoutent avec de plus en plus d'acuité les critères économiques (coût, délai, ...) exige de l'analyste la mise en œuvre d'outils et de méthodes de plus en plus fiables et répondant aussi efficacement que possible aux problèmes définis par un cahier de charges. Ainsi, toute approche d'analyse, vise en priorité à morceler la complexité apparente du système et à mettre en évidence son organisation interne. En effet, plus un système est vaste, plus le flux d'information véhiculé est important et plus il est difficilement maîtrisable. C'est pourquoi, toute approche envisagée, pour l'analyse d'un système, doit au moins posséder les caractéristiques suivantes :

- Généralité : Elle doit s'appliquer à tout type de système et ne doit requérir aucune hypothèse restrictive aussi bien au niveau des variables qu'au niveau des domaines d'application.
- Flexibilité : Elle doit s'adapter à toute modification pouvant surgir au niveau de l'organisation interne du système et l'évolution des données qui le caractérisent.
- Orientation : Elle doit faciliter la poursuite des phases ultérieures de l'analyse et donner une représentation aussi simplifiée que possible pour être accessible à tout utilisateur.
- Efficacité : Elle doit fournir des algorithmes faciles à mettre en œuvre, et rapides au niveau de leur capacité à formuler une solution.

Ainsi les méthodes de l'analyse structurale : l'analyse des données, la théorie de l'information, la théorie des graphes, la théorie des sous-ensembles flous, la théorie des jeux, les techniques d'intelligence artificielle s'avèrent plus adaptées pour l'appréhension des systèmes devenus de plus en plus complexes que les méthodes classiques d'analyse numérique, d'identification, etc.

L'application de ces approches suggère une méthodologie qui à partir d'opérations successives conduit à une structuration ou organisation du système analysé. Dans la hiérarchie épistémologique des systèmes de G.J. KLIR (KLIR 69, 75, 76, 77), plusieurs niveaux de connais-

sance sont mis en évidence. Entre ne rien connaître sur un système et le connaître complètement, il existe différents niveaux qui d'après KLIR s'organisent comme suit :

- Niveau 0 : Système source : C'est le niveau le plus bas, il correspond à une phase préparatoire. A ce niveau, sont définis l'ensemble des variables du système ainsi que leurs ensembles de modalités (ou états) respectives.
- Niveau 1 : Système donnée : le système est connu par une succession de modalités ou états pris par les variables, qui constituent le tableau initial des données que nous nous proposons d'analyser.
- Niveau 2 : Système générateur : Il est défini par un ensemble de relations génératrices invariantes ou fonctionnelles permettant la détermination du comportement de quelques variables en fonction de celui des autres.
- Niveau 3 : Système structure : A ce niveau le système est composé d'un ensemble de sous-systèmes générateurs qui interagissent entre eux et susceptibles de véhiculer une quantité importante d'information.

Les différents niveaux de la hiérarchie sont définis, toute approche d'analyse n'est autre qu'une suite de déplacements dans celui-ci. Ces déplacements peuvent être ascendants ou descendants selon le but visé par l'analyse. Mais avant de procéder à toute analyse, il convient de définir dans un premier temps le système, les variables mises en jeu, les notations et les structures algébriques nécessaires.

II - DESCRIPTION D'UN SYSTEME COMPLEXE

II.1 Description

Soit $\Sigma = \{X_i, i = 1, \dots, N\}$ l'ensemble des N variables et/ou paramètres importants, pertinents, pour la description du système. Il peut s'agir des variables d'entrée, de sortie, de commande, d'état ou bien de paramètres : constante de temps, amortissement, niveau de saturation, ... etc. Les variables retenues pour la description d'un système doivent mettre en évidence son fonctionnement et caractériser au mieux son observation. Selon la nature du système, certaines variables peuvent être accessibles, mesurables ou non, d'autres peuvent être inaccessibles et par conséquent non mesurables (nous discuterons la nature de ces variables au chapitre II). Concernant l'observabilité, M. PELEGRIN (PELEGRIN 81) dit à ce propos "la frontière du système doit être précisée et placée de telle façon que toutes les interactions avec l'extérieur soient observables". En effet, tracer une frontière revient à choisir une décomposition parmi l'ensemble des décompositions possibles du système. On considère par ailleurs, moyennant une

renumérotation donnée, que l'ensemble de ces variables primaires est ordonné (un ordre quelconque mais fixe). Notons également qu'aucune hypothèse algébrique n'est faite sur la nature de ces variables : elles peuvent être quantitatives ou qualitatives, logiques ou floues, à modalités structurées ou non.

- Soit $M = \{M_i, i = 1, \dots, N\}$ où M_i est l'ensemble (fini) des modalités prises par la variable $X_i, i \in (1, \dots, N)$ avec :

$$M_i = \{ \alpha_j^i, j = 1, \dots, m_i \}$$

Aucune hypothèse n'est faite sur l'ensemble de ces modalités. Dans certains cas, il convient de respecter la représentation naturelle des systèmes et éviter autant que possible de leur imposer des structures particulières (LERMAN 76). Nous verrons au paragraphe II.4 que le recodage d'une variable peut se traduire par une réduction du nombre de ses modalités. Par ailleurs, nous utiliserons également le mot état au lieu de modalité pour la description des systèmes dynamiques et particulièrement les systèmes séquentiels (appelés parfois machines à états). L'ensemble des modalités prises par chaque variable dépend du caractère statique ou dynamique de celui-ci.

• Echantillonnage et tableau de données :

Nous considérons une population d'apprentissage $\Omega = \{\omega_i, i = 1, \dots, L\}$ de taille L dont les variables sont observées aux instants d'échantillonnage : $t_0, t_0+\Delta t, t_0+2\Delta t, t_0+(L-1)\Delta t$ et génèrent le tableau initial de données (fig. I.1) que nous nous proposons d'analyser.

Dans ce tableau, $X_i(\omega_j)$ désigne :

- La modalité prise par la variable X_i pour le $j^{\text{ème}}$ échantillon dans le cas d'un système statique

$\begin{matrix} \Sigma \\ \Omega \end{matrix}$	X_1	X_i	X_N
ω_1				
⋮				
ω_j	$X_i(\omega_j)$		
⋮				
ω_L				

Fig. I.1 : Tableau initial de données

- Le vecteur de modalités prises par la variable X_i pour les p derniers échantillons

$\{\omega_j, \omega_{j-1}, \dots, \omega_{j+p-1}\}$ dans le cas d'un système dynamique d'ordre présumé p . Ainsi, les variables $X_i, i \in \{1, \dots, N\}$ sont vues comme des applications de Ω dans l'ensemble des modalités que l'on définit comme suit :

- Cas d'un système statique :

$$\begin{aligned} X_i : \Omega &\rightarrow M_i = \{ \alpha_k^i, k=1, \dots, m_i \} \quad \forall i \in \{1, \dots, N\} \\ \omega_j &\rightarrow X_i(\omega_j) \end{aligned} \quad (I.1)$$

- Cas d'un système dynamique :

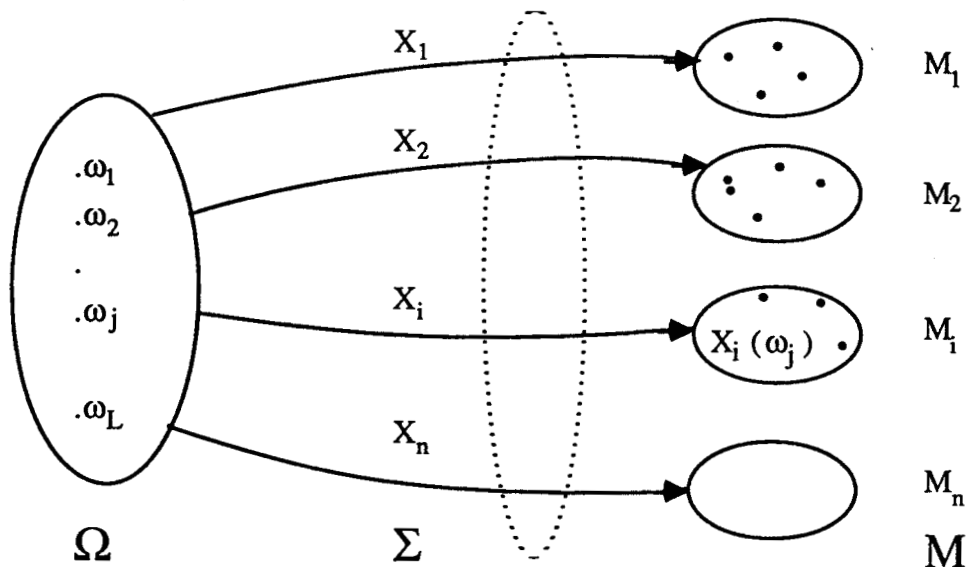
$$\begin{aligned} X_i : \Omega &\rightarrow \prod_{i=1}^P M_i \quad \forall i \in \{1, \dots, N\} \\ \omega_j &\rightarrow X_i(\omega_j) = \{ X_i(\omega_{j-k}), k = 0, \dots, p-1 \} \end{aligned} \quad (I.2)$$

Dans ce cas, chaque élément du tableau de données, est un vecteur à p composantes. Si on tient compte d'un tel vecteur au niveau des algorithmes, la représentation du système dynamique est alors identique à celle du système statique. Ainsi, à chaque fois que l'on parlera de tableaux de données, nous ne ferons aucune distinction entre le cas statique et le cas dynamique.

II.2 Variables et partitions

II.2.1 Le inf-demi treillis des partitions de Ω

Nous allons rappeler dans cette partie, quelques concepts de la théorie des ensembles qui seront utiles dans la suite de notre mémoire. En effet, un certain nombre d'opérations peuvent être effectuées sur les 3 ensembles Ω, M et Σ décrits précédemment et que l'on peut schématiser comme suit :



- Soit $\mathcal{P}(\Omega)$: l'ensemble des parties de Ω

- Une partition de Ω en k classes est un ensemble de k sous-ensembles :

$$C = \{C_i, i=1, \dots, k\} \text{ avec } C_i \in \mathcal{P}(\Omega) \text{ tel que :}$$

$$C_i \neq \emptyset, \forall i \in \{1, \dots, k\}$$

$$C_i \cap C_j = \emptyset, \forall i \neq j$$

$$\bigcup_i C_i = \Omega$$

i

- Soit $\mathcal{P}(\Omega) = \{C, C \text{ est une partition de } \Omega\}$. Dans cet ensemble, on peut établir un ordre à l'aide de la relation \leq (qui signifie plus fine ou égale à) définie comme suit :

$$\forall C, D \in \mathcal{P}(\Omega) : C \leq D \Leftrightarrow \forall C_i \in C, \exists D_j \in D \text{ tel que : } C_i \subset D_j \quad (\text{I.3})$$

$$\text{Autrement dit : } \forall D_j \in D, D_j = \bigcup_i C_i, i \text{ tel que : } C_i \subset D_j$$

Ainsi D détermine une partition des classes de C ou bien D couvre C .

La relation " \leq " est un ordre partiel (LERMAN 81). Cette relation d'ordre induit sur l'ensemble $\mathcal{P}(\Omega)$ une structure de treillis. Tout couple de partitions C et D admet un plus petit majorant (noté $C \vee D$) et un plus grand minorant (noté $C \wedge D$) définis comme suit :

$$(C \wedge D) \leq C \text{ et } (C \wedge D) \leq D, \text{ et } \forall E \text{ tel que } E \leq C \text{ et } E \leq D \text{ alors } E \leq (C \wedge D) \quad (\text{I.4})$$

$$(C \vee D) \geq C \text{ et } (C \vee D) \geq D, \text{ et } \forall E \text{ tel que } E \geq C \text{ et } E \geq D \text{ alors } E \geq (C \vee D) \quad (\text{I.5})$$

Ainsi, $\mathcal{P}(\Omega)$ muni de l'opération \wedge telle que :

$$\begin{aligned} \wedge : \mathbb{P}(\Omega) \times \mathbb{P}(\Omega) &\rightarrow \mathbb{P}(\Omega) \\ (C,D) &\rightarrow C \wedge D \text{ défini par (I.4)} \end{aligned}$$

vérifie les propriétés d'idempotence, de commutativité et d'associativité et constitue un inf-demi-treillis (TORO 82, LERMAN 81).

Chaque variable X_i (resp sous-ensemble de variables S), définit une partition de Ω notée P_{X_i} (resp P_S) de la façon suivante :

$$P_{X_i} = \{ X^{-1}(\alpha) : \alpha \in M_i \} = \{ \{ \omega : X(\omega) = \alpha \} : \alpha \in M_i \}$$

Ceci signifie que la partition de Ω est obtenue par regroupement dans une même classe de tous les ω auxquels X_i attribue la même modalité α . Il faut noter également que P_{X_i} ne dépend pas des modalités en elles-mêmes, mais seulement du fait qu'elles sont distinctes (DUSSAUCHOY 80).

A travers l'association $X_i \in \Sigma \rightarrow P_{X_i} \in \mathbb{P}(\Omega)$ on peut dire que : $\Sigma \subseteq \mathbb{P}(\Omega)$.

Cette possibilité d'associer de façon univoque à chaque variable une partition induit un ordre et une équivalence entre variables. Ainsi, si on définit 2 variables sur Ω de la façon suivante :

$$\begin{array}{ll} X : \Omega \rightarrow M_X & Y : \Omega \rightarrow M_Y \\ \omega \rightarrow X(\omega) & \omega \rightarrow Y(\omega) \end{array}$$

On dira que :

$$i) \quad X \text{ est équivalente à } Y \Leftrightarrow P_X = P_Y \quad (I.6)$$

Cela veut dire qu'il est possible de trouver une fonction $f : M_X \rightarrow M_Y$ bijective telle que $Y = f \circ X$ et $X = f^{-1} \circ Y$, il s'agit tout simplement d'un recodage

$$ii) \quad X \text{ plus fine que } Y \Leftrightarrow P_X \leq P_Y \Leftrightarrow \exists f : M_X \rightarrow M_Y \text{ telle que : } Y = f \circ X \quad (I.7)$$

Compte tenu du fait que les concepts de la théorie de l'information ne distinguent pas entre variables équivalentes, on peut interpréter $\mathbb{P}(\Omega)$ comme l'ensemble de toutes les variables définies sur Ω .

II.2.2 Le sup-demi treillis des variables vectorielles

Soit $\mathcal{P}(\Sigma) = \{S : S \subseteq \Sigma\}$: l'ensemble des parties de Σ

$\mathcal{P}(\Sigma)$ n'est autre que l'ensemble des variables vectorielles dont les composantes sont des variables de Σ . Ainsi, si on note :

$$- S = \{X_{i_1}, X_{i_2}, \dots, X_{i_k}\} \in \mathcal{P}(\Sigma)$$

$$- J_S = \{i, X_i \in S\} = \{i_1, i_2, \dots, i_k\} : \text{l'ensemble des sous indices de } X_i \in S, \text{ avec :}$$

$$M_S = \prod_{i \in J_S} M_i = M_{i_1} \cdot M_{i_2} \cdot \dots \cdot M_{i_k}$$

la variable vectorielle S est alors définie comme suit :

$$\begin{aligned} S : \Omega &\rightarrow M_S \\ \omega &\rightarrow S(\omega) = \{X_{ij}, j = 1, \dots, k\} \end{aligned}$$

De ce fait, chaque élément $S \in \mathcal{P}(\Sigma)$, représente toutes les variables vectorielles qui lui sont équivalentes par permutation de ses composantes. Par ailleurs, du fait que $\mathcal{P}(\Sigma)$ est l'ensemble des variables définies sur Ω , où chacune d'elles induit une partition de $\mathcal{P}(\Omega)$, on dira alors que : $\mathcal{P}(\Sigma) \subseteq \mathcal{P}(\Omega)$. Dans $\mathcal{P}(\Sigma)$, la relation d'inclusion \subseteq possède les propriétés de réflexivité, d'anti-symétrie et de transitivité et définit par conséquent un ordre partiel tel que à tout couple $A, B \in \mathcal{P}(\Sigma)$ est associé un infimum (correspondant à $A \cap B$) et un supremum (correspondant à $A \cup B$) :

$$(A \cap B) \subseteq A \text{ et } (A \cap B) \subseteq B, \text{ et } \forall D \text{ tel que : } D \subseteq A \text{ et } D \subseteq B \Rightarrow D \subseteq (A \cap B) \quad (\text{I.7})$$

$$(A \cup B) \supseteq A \text{ et } (A \cup B) \supseteq B, \text{ et } \forall D \text{ tel que : } D \supseteq A \text{ et } D \supseteq B \Rightarrow D \supseteq (A \cup B) \quad (\text{I.8})$$

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

$$\begin{aligned} U : \mathcal{P}(\Sigma) \cdot \mathcal{P}(\Sigma) &\rightarrow \mathcal{P}(\Sigma) \\ (A, B) &\rightarrow A \cup B \text{ défini par (I.8)} \end{aligned}$$

vérifie les propriétés d'idempotence, de commutativité et d'associativité et constitue un sup-demi-treillis.

Si on considère deux sous-ensembles de variables $A, B \in \mathcal{P}(\Sigma)$ et $P_A, P_B ; P_{A \cup B}$ les partitions qu'ils induisent sur Ω , on a alors les deux propriétés suivantes :

$$i) A \subseteq B \Rightarrow P_A \geq P_B \quad (I.9)$$

$$ii) P_{A \cup B} = P_A \wedge P_B \quad (I.10)$$

Ces deux propriétés mettent en évidence la relation qui existe entre $P(\Omega)$ et $\mathcal{P}(\Sigma)$.

II.3 Estimation des probabilités

Sous sa forme générale (PARZEN 60, DACHUNA 82), une mesure de probabilité sur un ensemble fini Ω est une fonction :

$$\begin{aligned} \text{Pr} : \mathcal{P}(\Omega) &\rightarrow \mathbb{R} \\ A &\rightarrow \text{Pr}(A) \end{aligned} \quad (I.11)$$

qui vérifie les propriétés suivantes :

- i) $\forall A \in \mathcal{P}(\Omega) : \text{Pr}(A) \geq 0$
- ii) $\forall i=1, \dots, k \quad A_i \in \mathcal{P}(\Omega)$ et $i \neq j, \quad A_i \cap A_j = \emptyset : \text{Pr}(\bigcup_i A_i) = \sum_i \text{Pr}(A_i)$
- iii) $P(\Omega) = 1$

Le caractère aléatoire des variables observées, est dû soit à la nature du système lui-même ou bien résulte de la manière dont il a été excité.

Ainsi, si $S : \Omega \rightarrow M_S$ est une variable quelconque définie sur Ω , on définit pour toute modalité $\alpha \in M_S$ une probabilité d'occurrence notée : $\text{Pr}(S = \alpha)$.

Sous certaines conditions concernant le système (ergodicité, stationnarité), la taille de l'échantillon et la période d'échantillonnage, de telles probabilités peuvent être estimées au moyen des fréquences relatives évaluées sur Ω et s'écrivent :

$$\text{Pr}(S=\alpha) = \frac{|\{ \omega : S(\omega) = \alpha \}|}{|\Omega|} \quad (I.12)$$

dans le cas d'un système statique, avec $|\Omega| = L$ et S une variable statique

$$\text{Pr}(S=\alpha) = \frac{|\{ \omega : S(\omega) = \alpha \}|}{|\Omega|} \quad (I.13)$$

dans le cas d'un système dynamique, avec $|\Omega| = L-p+1$ et S une variable dynamique.

L'estimation de ces probabilités nous conduit à faire les remarques suivantes.

a - Le choix de $|\Omega|$ joue un rôle important quant à la représentativité et la validité des résultats d'une analyse. En effet, un échantillon de faible cardinal ne peut conduire qu'à des résultats provisoires, c'est pourquoi, celui-ci doit être suffisamment élevé pour une estimation correcte des probabilités. CONANT (CONANT 72) propose une limite inférieure :

$$|\Omega| \geq \prod_{i=1}^N (|M_i| - 1)$$

D'un point de vue opérationnel, dans la mesure où l'on ne suppose aucune distribution particulière (normale, exponentielle, ...), le choix du nombre d'échantillons dépend essentiellement de la nature du système et de l'expérience du praticien. Nous montrerons, dans le cas d'une station d'épuration des eaux (chapitre II), que le nombre d'échantillons à retenir doit tenir compte essentiellement des modes de fonctionnement (faibles, moyennes et fortes charges) du système.

b - L'évolution des systèmes fait intervenir la notion de temps, il faudrait donc en toute rigueur que l'hypothèse de stationnarité et d'ergodicité soit vérifiée. Par ailleurs, le système ne peut au cours du temps, évoluer d'un état à un autre sans passer au moins une fois par tous les états intermédiaires (OSWALD 86), par conséquent le choix d'une période d'échantillonnage T_1 optimale doit être égale à la durée minimale T_0 des états du système. Dans le cas où $T_0 = T_1$: l'échantillonnage fournit sans redondance ni lacune l'information nécessaire pour décrire complètement l'évolution du système (SHANNON 48) (OSWALD 86).

c - L'analyse du système, nécessite dans un premier temps, l'estimation des probabilités multidimensionnelles. Or un tel calcul, à l'aide des fréquences relatives, est simple mais le nombre d'opérations nécessaires à sa mise en œuvre croît de façon exponentielle avec le nombre de variables et/ou le nombre de leurs modalités respectives. D'où la nécessité de recourir à un recodage des variables pour limiter le nombre de leurs modalités et par conséquent le temps de calcul qui en résulte.

II.4 Recodage des variables

Le problème de l'estimation des probabilités, a été largement étudié dans la littérature [WAGNER 75, YOUNG 78, DUBUISSON 80, MORGERA 81, OSWALD 86]. Pour faciliter cette opération, lorsque on procède à l'analyse du système, on est amené parfois à regrouper ses variables et/ou ses modalités. Ceci est d'autant plus vrai que le système est de dimension impor-

tante. Les conditions classiques sont telles que le regroupement s'effectue :

- a) sans détérioration de la structure algébrique de la variable ;
- b) avec un minimum de perte d'information au sens général du terme ;
- c) en rendant plus ou moins équiprobable les différentes modalités [CONANT 76].

Or toute perte d'information dans ce cas, se traduit par une diminution de la capacité d'observation de la variable en question [DUFOUR 76]. Par ailleurs, équiprobabiliser les états ou modalités d'une variable revient à classer les individus selon des critères de masses et non pas pour des critères de description des données elles-mêmes, ceci conduit naturellement à une classification arbitraire. Il convient de tenir compte également de la nature des variables et l'information qu'elles recèlent (valeurs caractéristiques, domaine de variation, variable à expliquer, variable explicative) [JAMBU 78], ainsi que de l'absence ou la présence du bruit sur ces variables.

II.4.1 Position du problème

Soit $\mathbb{P}_k(X) = \{F_1, F_2, \dots, F_k\}$ une partition de X en k classes induite par les données sur Ω . Il est alors possible de générer une partition de X en j classes $\mathbb{P}_j(X) = \{C_1, C_2, \dots, C_j\}$ plus grosse que $\mathbb{P}_k(X)$ telle que chaque classe $C_i \in \mathbb{P}_j(X)$ soit l'union des classes F_i correspondant au regroupement des modalités de X . Moyennant une renumérotation donnée, on peut écrire :

$$\mathbb{P}_k(X) = \{F_1, F_2, \dots, F_a, F_{a+1}, \dots, F_b, F_{b+1}, \dots, \dots, F_u, \dots, F_k\} \quad (I.14)$$

$$\mathbb{P}_j(X) = \{ C_1, C_2, \dots, C_i, \dots, C_j \}$$

Ainsi $\mathbb{P}_j(X)$ couvre ou détermine une partition des classes de $\mathbb{P}_k(X)$. Il s'agit alors de trouver parmi toutes les partitions moins fines que $\mathbb{P}_k(X)$ (partition "naturelle" associée à X), celle qui optimise un critère donné. Si on note $\mathbb{P}_j^*(X)$ une telle partition, celle-ci est telle que :

$$\mathbb{P}_j^*(X) \succ \mathbb{P}_j(X) \quad \forall \mathbb{P}_j(X) \in \mathbb{P}_j(\Omega) \quad (I.15)$$

avec " \succ " qui signifie : "préférer à" et $\mathbb{P}_j(\Omega)$: l'ensemble des partitions de Ω en j classes ($j \leq k$).

II.4.1.1 Cas des variables ordinales

La variable X étant ordinale, il est nécessaire de tenir compte de la contrainte supplémentaire concernant la relation d'ordre existant entre les classes $F_i \in \mathbb{P}_k(X)$. En effet, une telle relation doit être conservée lors de la reconstitution des classes $C_i \in \mathbb{P}_j(X)$ auxquelles on doit imposer la restriction suivante :

Si on note \mathfrak{R} la relation d'ordre existante, alors :

- $\forall F_i, F_{i+1} \in \mathbb{P}_k(X)$ on a : $F_i \mathfrak{R} F_{i+1}$ d'où :
- $\forall C_i, C_{i+1} \in \mathbb{P}_j^*(X)$ on doit avoir : $C_i \mathfrak{R} C_{i+1}$ ceci entraîne naturellement

que :

$$C_i = \bigcup_{r: F_r \subset C_i} F_r \quad \text{tel que } F_r \subset C_i, F_r \mathfrak{R} C_{i+1} \text{ et } C_{i-1} \mathfrak{R} F_r \quad (\text{I.16})$$

Un tel problème peut être formulé comme un problème de classification sous contrainte d'ordre [LECHEVALIER 76].

II.4.1.2 Cas des variables quantitatives

Dans ce cas, X prend ses modalités dans \mathbb{R} et plus précisément dans un ou plusieurs segments $[a, b[$. On peut alors associer à chaque classe $C_i \in \mathbb{P}_j^*(X)$ un intervalle V_i, V_{i+1} . Ainsi, le rangement en classes des $\omega_i \in \Omega$ permettant de générer $\mathbb{P}_j(X)$ s'énonce alors :

$$\forall \omega \in \Omega, \omega \in C_i \Leftrightarrow X(\omega) \in [V_i, V_{i+1}[, \text{ avec } [V_i, V_{i+1}[\subset [a, b[\quad (\text{I.17})$$

En effet, établir $\mathbb{P}_j^*(X)$ revient à découper de façon optimale le segment $[a, b[$ en j intervalles, tels que :

- $\bigcup_{i=1}^j [V_i, V_{i+1}[= [a, b[$
- $[V_i, V_{i+1}[\cap [V_k, V_{k+1}[= \emptyset, \quad k \neq i$

Autrement dit, ceci revient à déterminer les $(j-1)$ valeurs séparatrices des classes étant donné que $V_1 = a$ et $V_{j+1} = b$

II.4.2 Caractérisation de $\mathbb{P}_j^*(x)$:

La mesure de la qualité du recodage de la variable X dépend de la nature de la variable elle-même et également des critères retenus pour juger la qualité de $\mathbb{P}_j(X)$. Pour caractériser une telle partition, deux critères sont mis en jeu : critères portant sur les caractéristiques des classes et les critères dépendant du type de variables (variables à expliquer, variables explicatives).

II.4.2.1 Critères relatifs aux caractéristiques des classes :

a) Classes équidistantes : Le découpage du domaine des valeurs de X en j intervalles de même longueur, permet d'établir la partition $\mathcal{P}_j^*(X)$ en j classes équidistantes. Ainsi, si on note (a,b) les bornes inférieure et supérieure de ce domaine, la longueur unitaire associée à chaque classe C_i est donnée par :

$$L = (b - a)/j \quad (\text{I.18})$$

La relation de récurrence suivante détermine les valeurs séparatrices des classes :

$$i = 1, \dots, j \quad : \quad V_{i+1} = V_i + L \quad \text{avec } V_1 = a$$

b) Classes équiprobables : Dans ce cas, il s'agit de trouver la partition $\mathcal{P}_j^*(X)$ qui présente des classes homogènes (de même "poids"), contenant le même nombre d'observations de Ω .

Si on note $f(X)$ la densité de probabilité de X sur Ω telle que :

$$f(X) > 0 \quad \text{et} \quad \int_a^b f(X) dX = 1$$

Les classes séparatrices de $\mathcal{P}_j^*(X)$ seront alors équiprobables si elles vérifient l'expression suivante :

$$i = 1, \dots, j \quad : \quad \int_{V_i}^{V_{i+1}} f(X) dX = \frac{1}{j} \quad , \quad \text{avec } V_1 = a \quad (\text{I.19})$$

On sait que la partition ainsi obtenue (équiprobabilité des classes) est celle qui permet d'extraire le maximum de l'information (au sens de SHANNON) sur la variable X . Nous reviendrons sur cet aspect au paragraphe III.

Les deux critères ainsi définis, ne suffisent pas à eux seuls à déterminer la valeur optimale de " j ". Ainsi le nombre de modalités à retenir doit être choisi également en fonction des connaissances à priori sur le système.

II.4.2.2 Critères dépendant du type de variables

Il s'agit de tenir compte de la nature des variables en absence ou en présence du bruit : variables à expliquer, variables explicatives primaires, variables explicatives secondaires, variables internes, ... etc. Nous aborderons ce problème dans le chapitre II consacré à la modélisation.

Après cette description des systèmes complexes, nous allons présenter certains outils permettant l'analyse de tels systèmes, sur lesquels on ne possède aucune information à priori. A ce stade, les seules connaissances dont dispose généralement le praticien se limitent aux variables qu'il croit pertinentes pour la description du système, et à partir desquelles un tableau de données est généré. Par ailleurs, guidé par son "bon sens" et son expérience, le praticien peut émettre quelques idées sur les relations entre certaines variables du système, peut être également amené à distinguer concrètement ou directement des ensembles d'éléments du système, connectés faiblement ou fortement par des flux de matière, d'information, d'énergie, ... etc. Celui-ci peut effectivement dégager quelques grandes lignes de la structure du système et uniquement celles-ci. En revanche, il se révèle incapable de procéder à des analyses plus fines du système [DUFOR 79]. De telles analyses peuvent être abordées, à l'aide des méthodes géométriques de "l'analyse des données" et des méthodes issues des concepts de la théorie de l'information. Nous présenterons ci-dessous l'apport de ces outils pour l'analyse d'un système complexe.

III - OUTILS UTILISES POUR L'ANALYSE D'UN SYSTEME COMPLEXE

L'analyse structurale des systèmes est orientée essentiellement vers les buts suivants :

- identifier les sous-systèmes éventuels (problème de décomposition) ;
- mettre à jour les redondances, c'est-à-dire l'organisation interne du système (problème d'explication) [WALLISER 77].

Dans ce sens, les techniques d'Analyse des données, permettent dans certains cas la conduite d'une telle analyse.

III.1 Les outils géométriques de l'analyse de données

L'analyse de données est toujours présentée comme un outil d'analyse de systèmes complexes dont on connaît peu et même parfois pas du tout la structure, et sur lesquels on a pu faire un grand nombre de relevés expérimentaux [GENTIL 84]. Il existe actuellement, plusieurs méthodes d'analyse de données qui se distinguent les unes des autres par le type de variables qu'elles traitent et par le choix de la métrique utilisée pour évaluer les couplages et les inerties du système. De telles méthodes peuvent se ranger en trois catégories :

- celles qui servent à "visualiser" les données, c'est-à-dire les présenter d'une façon qui facilite la décision intuitive ;
- celles qui servent à "expliquer", c'est-à-dire permettre d'estimer la valeur d'une variable en fonction d'autres appelées variables explicatives [LABORDERE 77] ;
- celles qui servent à structurer ou organiser les différentes composantes du système.

Dans tous les cas, le choix d'une méthode se justifie par l'objectif visé par l'analyse et la nature des données à traiter.

Nous présenterons ci-dessous, quelques unes des principales méthodes d'analyse des données qui peuvent être utilisées également comme outil d'analyse structurale d'un système complexe en vue de dégager ses caractéristiques importantes. Nous aborderons successivement l'analyse en composantes principales, l'analyse des correspondances (appelées généralement méthodes factorielles), la régression linéaire, l'analyse canonique et quelques techniques de décomposition. Sans vouloir nous étendre en de longs développements qui dépasseraient le cadre de ce travail, signalons toutefois que ces méthodes ont fait l'objet d'un grand nombre de travaux, dans les domaines les plus variés [BERTIER 75] [CAILLIEZ 76] [BENZECRI 80] [VOLLE 82].

III.1.1 Analyse en composantes principales

On dispose d'un tableau initial de données de dimension $N.L.$ qui peuvent être représentées par un nuage de N points dans l'espace \mathcal{R}^L des variables qui sont essentiellement des variables quantitatives. La représentation de l'ensemble des points ne peut se faire, dans un espace de dimension importante, on cherche alors une visualisation dans un espace de dimension réduite, facilitant ainsi l'interprétation de la quantité d'information présente dans ces données. Il s'agit de trouver un espace \mathcal{R}^K ($K < L$) occasionnant le minimum de perte d'information (l'information est mesurée par l'inertie du système). On procède alors comme suit :

- on cherche une droite telle que l'inertie du nuage par rapport à cette droite soit minimum. Elle constitue le premier axe principal, le nuage est alors représenté par la projection des points sur cet axe ;
- l'opération est répétée dans \mathcal{R}^{L-1} , on obtient alors sur le deuxième axe et un nouveau nuage projeté dans \mathcal{R}^{L-2} .

Les facteurs principaux sont des nouvelles variables obtenues à partir des variables initiales de la manière suivante :

$$Y_1 = \sum_{i=1}^N Y_i^1 X_i \quad : \text{ telle que la variance de } Y_1 \text{ soit maximum} \quad (I.20)$$

$$Y_2 = \sum_{i=1}^N Y_i^2 X_i$$

Y_2 orthonormale et non corrélée à Y_1 et telle que la variance de Y_2 soit maximum et ainsi de suite.

Les facteurs principaux et les axes principaux sont liés. Les facteurs principaux sont les variables selon lesquelles la dispersion de l'ensemble d'échantillonnage est maximale. On peut vérifier également que ces facteurs principaux constituent des vecteurs propres de la matrice de variance-covariance des variables du système.

III.1.2 Analyse factorielle des correspondances

Cette méthode s'applique essentiellement à des tableaux de contingence croisant deux ensembles de modalités. C'est une méthode basée sur l'utilisation de la métrique du χ^2 , et permet de décrire la dépendance entre variables qualitatives. Le choix d'une telle distance est lié principalement à la notion d'équivalence distributionnelle. En effet, l'agrégation de deux lignes ou deux colonnes des tableaux n'occasionne aucune perte d'information si celles-ci présentent les mêmes profils et ne modifie pas par conséquent les résultats de l'analyse. Signalons, toutefois que cette méthode est la plus puissante des méthodes de l'analyse des données et fournit des résultats clairement interprétables [VOLLE 76, BENZECRI 80]. Nous comparerons cette méthode avec une approche que nous développerons au chapitre III.

III.1.3 Régression linéaire

On considère deux sous-ensembles de variables X et $Y \in \mathcal{P}(\Sigma)$ entre lesquels on désire analyser les relations. On note $Y_i, i \in \{1, \dots, N\}$ les variables à expliquer et

$$X_i^j, i \in \{1, \dots, N\}, j \in \{1, \dots, M\}$$

les variables explicatives. Sous l'hypothèse de l'existence de relations linéaires entre elles, on cherche $(N+1)$ scalaires $a_i ; i = 0, \dots, N$ tel que :

$$Y_i - (a_0 + a_1 X_i^1 + a_2 X_i^2 + \dots + a_N X_i^M) \quad (I.21)$$

est minimum $\forall i \in \{1, \dots, N\}$

D'une manière générale, ceci peut s'écrire avec une notation vectorielle comme suit :

$$\text{trouver } a \in \mathbb{R}^{N+1} \text{ tel que } \| Y - X_a \| \text{ est minimum.}$$

La qualité de la représentation de Y par X_a peut être mesurée à l'aide de l'indice :

$$\frac{\| X_a \|}{\| Y \|}$$

III.1.4 Analyse canonique

A partir du tableau initial de données, on extrait deux catégories de variables $Y^T = (Y_1, Y_2, \dots, Y_p)$ et $X^T = (X_1, X_2, \dots, X_q)$ entre lesquelles il s'agit d'analyser les relations. On cherche une forme linéaire des Y_j :

$$\alpha_i = (Y, a_i) = \sum_{j=1}^p a_i^j Y_j$$

(espace des formes linéaires de dimension p) et une forme linéaire des X_k :

$$\beta_i = (X, b_i) = \sum_{k=1}^q b_i^k X_k$$

(espace des formes linéaires de dimension q) tels que α_i et β_i ont une corrélation maximum.

La procédure de recherche des α_i et β_i peut être formulée comme suit :

- trouver $a_1 \in \mathbb{R}^p$ et $b_1 \in \mathbb{R}^q$ tels que :

- la corrélation entre α_1 et β_1 est maximum sous la contrainte :
- $\text{var}(\alpha_1) = 1$ et $\text{var}(\beta_1) = 1$

- trouver $a_2 \in \mathbb{R}^p$ et $b_2 \in \mathbb{R}^q$ vérifiant les conditions précédentes ainsi que les conditions de non corrélation :

$$\text{Cor}(\alpha_1, \alpha_2) = 0 \text{ et } \text{Cor}(\beta_1, \beta_2) = 0 \quad (\text{I.22})$$

Une telle procédure peut être répétée au plus p fois ($p \leq q$), les α_i et β_i sont appelés variables canoniques associées aux variables Y_j [resp aux variables X_k].

La plupart des méthodes précédemment décrites utilisent l'hypothèse de linéarité des re-

lations entre variables. Par ailleurs, chaque méthode est en général adaptée à un type de système. En effet, selon la nature du tableau de données (homogène, hétérogène, de contingence, binaire, préférence, ...), la nature des variables mises en jeu (quantitatives, qualitatives, ...) et l'objectif escompté, il convient d'appliquer la méthode qui s'avère la mieux appropriée.

III.2 Les outils de la théorie de l'information

III.2.1 Introduction

La théorie de l'information (T.I.) est née en 1948 avec les travaux de C.E. SHANNON [SHANNON 48] et de N. WIENER qui ont connu d'emblée un retentissement considérable. Conçue initialement pour l'étude de la transmission des messages émis par une source à un récepteur à travers un canal soumis au bruit, celle-ci a connu un développement rapide et fécond [SLEPIAN 74]. De nombreux travaux ont été consacrés à cette théorie, qui ne sont en définitive qu'une exploitation et extension des travaux proposés par SHANNON, HARTLEY, WIENER et NYQUIST. Cette théorie a subi plusieurs mutations, de grandes voies se sont dégagées principalement dans les trois domaines suivants :

a) Communication : Dans ce cas, on étudie les phénomènes liés à l'acheminement des signaux, aux organes de transmission, aux fonctions de modulation, de détection, ... etc. Pour une étude bibliographique, voir [WYNER 81], [OSWALD 86].

b) Mathématique : La T.I. est vue sous un angle plus abstrait. Dans le cadre de la théorie de la mesure, de l'information relativiste, plusieurs généralisations ont été proposées soit au niveau purement mathématique ou bien pour des applications particulières. Nous citons parmi ces auteurs : KAMPE DE FERIET, BENVENUTI, B. FORTE sur l'information généralisée ; C. LANGRAND, J. LOSFELD, A. DUSSAUCHOY, J.P. BARTHELEMY, S. GUIASU pour d'autres types d'informations que celles de SHANNON et enfin G. JUMARIE pour l'information relativiste. Pour une illustration de la liaison entre les diverses mesures d'information, voir les colloques du C.N.R.S. sur la théorie de l'information et ses applications (CACHAN, juillet 77), A.F.C.E.T. (Versailles, novembre 77), voir aussi [GUIASU 77] [SALLATIN 79] [PICARD 78].

c) Analyse des systèmes : Elle a pour objet l'étude des flux d'informations échangés entre les différentes composantes du système. C'est dans cette optique que nous nous plaçons pour présenter des résultats relatifs à l'analyse des systèmes dans l'esprit de l'Automatique.

Si la littérature consacrée aux deux premiers domaines a été et continue d'être abondante, elle l'est par contre moins et reste assez limitée dans le domaine qui nous intéresse : l'analyse structurale des systèmes. Les premiers travaux dans ce domaine sont dûs essentiellement du côté américain à W.R. ASHBY [56, 65, 69], R.C. CONANT [69, 70, 76], G. KLIR [1969,

75, 76, 77] et à M. RICHTIN [1975], J. DUFOUR [1979], DUSSAUCHOY [1980], V. TORO [1980] du côté français. Ainsi, dans la plupart de ces travaux, on met en évidence les notions relatives aux lois de transmission et d'évolution de l'information dans un système, on étudie les relations qui existent entre la disponibilité de l'information et la capacité de contrôle et de régulation d'un système. On distingue également les aspects de sélection, de blocage, de transmission interne et externe des quantités d'informations créées ou relayées par le système.

Toutes ces démarches reposent sur des concepts de mesure d'information et d'entropie.

III.2.2 Mesure d'information et entropie

Soit A un événement, sa réalisation apporte une quantité d'information et on note $I_\phi(A)$ ce nombre [AGGRWAL 74] tel que :

$$I_\phi(A) : \phi_0(P(A)) - \phi_0(P(\Omega)) \text{ où :}$$

- Ω : l'ensemble des événements élémentaires ω
- $P(A)$: la probabilité d'apparition de l'événement A
- ϕ : une application de $[0,1]$ dans \mathbb{R}^+ telle que $\phi(1) \geq 0$

Par ailleurs, on sait que la réalisation d'un événement peu probable apporte beaucoup d'information alors que la réalisation d'un événement très probable (ou presque certain) apporte peu d'information. C'est pourquoi, une mesure d'information doit être une mesure décroissante de probabilité. Toujours selon AGGRWAL [1974], $I_\phi(A)$ vérifie les axiomes fondamentaux de KAMPE DE FERRIET et FORTE. L'information $H_\phi(X)$ fournie par la partition

$X = \{A_1, \dots, A_n\}$ est définie par :

$$H_\phi(x) = \sum_{i=1}^n P(A_i) I_\phi(P(A_i)) = \sum_{i=1}^n P(A_i) \phi(P(A_i)) - \phi(1) = \sum_i P_i \phi_i - \phi(1)$$

• si $\phi(t) = \left(\frac{1}{t} - 1\right) \delta(t - P_M)$ où $\begin{cases} P_M = \max_i P_i \\ \delta(t - P_M) : \text{l'impulsion de Dirac au point } P_M \end{cases}$

$H_\phi(X)$ s'écrit alors :

$$H_\phi(x) = \sum_i P_i \left(\frac{1}{P_i} - 1\right) \delta(P_i - P_M) = P_M \left(\frac{1}{P_M} - 1\right) = 1 - P_M$$

c'est l'entropie du Max que nous noterons :

$$H_M(x) = 1 - \max_i P_i = 1 - P_M \quad (I.23)$$

• si $\phi(t) = -\text{Log } t$, $H_\phi(x)$ s'écrit :

$$H_\phi(x) = \sum_{i=1}^n -P_i \log P_i \quad (I.24)$$

c'est l'entropie de SHANNON

III.2.3 Une autre approche pour la définition de l'entropie de SHANNON d'une variable discrète :

Nous allons commencer ce paragraphe, par un exemple intuitif, sur la mesure de la quantité d'information, que nous empruntons à J. OSWALD [1986]. Soit un système (S), capable de prendre L états différents, non équiprobables. Pour une configuration donnée correspondant à N situations successives de (S), on distinguera n_i fois l'état S_i , $i = 1, \dots, L$. Le nombre de situations différentes $P(N)$, d'ordre N s'écrit :

$$P(N) = \frac{N!}{n_1! n_2! \dots n_L!} \quad \text{avec} \quad \sum_{i=1}^L n_i = N$$

Or la définition de la quantité d'information conduit à évaluer le nombre de configurations possibles par $\text{Log}_2 P(N)$ [en désignant par Log_2 le logarithme à base 2]. Par ailleurs si N est suffisamment élevé pour que les n_i soient eux-mêmes des nombres élevés et que l'approximation de STIRLING soit valable, en remplaçant $\text{Log } N!$ par $N \text{Log } N - N$ et $\text{Log } n_i! = n_i \text{Log } n_i - n_i$ et en tenant compte des relations précédentes, on peut alors écrire :

$$\frac{1}{N} \text{Log } P(N) = - \sum_{i=1}^L \frac{n_i}{N} \text{Log } \frac{n_i}{N} \quad (I.25)$$

n_i/N est la fréquence d'apparition de l'état S_i pour un ordre N donné. Lorsque celui-ci devient suffisamment élevé, la limite n_i/N tend par définition vers la probabilité d'apparition de l'état S_i que nous noterons P_i .

La quantité :

$$\sum_1^L P_i \text{Log } P_i^{-1} = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Log } P(N) = H \quad (\text{I.26})$$

est appelée par OSWALD : densité diacritique du système à états indépendants quelconques, qui représente également son entropie. Elle mesure son indétermination à priori. Il est clair que les renseignements que l'on extrait de (S) seront d'autant plus importants que son incertitude à priori est élevée. Par ailleurs, compte tenu de la fonction $P \text{Log } P^{-1}$, les états qui apportent la plus forte contribution à H, sont ceux qui présentent des probabilités, ni trop faibles, ni trop élevées puisque le maximum d'une telle fonction est atteint pour $P_i = e^{-1}$.

• Définition formelle de l'entropie :

Dans la mesure où chaque variable du système est défini sur Ω , l'entropie peut être vue comme une fonction qui s'applique aussi bien aux variables ou groupes de variables de Σ qu'aux partitions de Ω .

Soit $S : \Omega \rightarrow M_S$ une variable discrète quelconque définie sur Ω , et

$P(S = \alpha) = P(S^{-1}(\alpha)) = P(\{\omega : S(\omega) = \alpha\})$: probabilité de $S = \alpha$, l'entropie de S peut être définie comme une application de $\mathcal{P}(\Sigma)$ dans \mathbb{R} , ou comme une application de $\mathbb{P}(\Omega)$ dans \mathbb{R} .

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

$$S \longrightarrow H(S) = - \sum_{\alpha} P(S = \alpha) \text{Log}(S = \alpha) \quad (\text{I.27})$$

• $H : \mathbb{P}(\Omega) \longrightarrow \mathbb{R}$

$$P_S \longrightarrow H(P_S) = - \sum_{\alpha} P(S_{\alpha}) \text{Log}(S_{\alpha}) \quad (\text{I.28})$$

avec $H(\phi) = 0$

Deux variables équivalentes donnent naissance à la même partition de Ω et possèdent la même entropie.

• Unité de l'entropie :

L'unité dépend essentiellement de la base choisie pour le logarithme. Ainsi dans le cas

d'une base 2, l'unité est alors le bit (binary digit) et correspond à l'entropie d'une variable prenant deux modalités, chacune avec une probabilité 1/2. Dans le cas d'un logarithme népérien (base e), l'unité de H est exprimée en Nat (Natural unit). Certains auteurs préconisent l'utilisation du terme shannon pour la première unité et le logon pour la seconde unité, ainsi :

$$\begin{aligned} 1 \text{ shannon} &\cong 0,69 \text{ logon} \\ 1 \text{ logon} &\cong 1,44 \text{ shannon} \end{aligned}$$

Désormais, nous utiliserons le symbole Log pour désigner le logarithme à base 2.

III.2.4 Les propriétés de l'entropie de SHANNON

L'entropie de SHANNON possède un certain nombre de propriétés qui rendent son utilisation intéressante dans l'analyse des systèmes.

$$- \forall S \in \mathcal{P}(\Sigma) : H(S) \geq 0, H(S) = 0 \Leftrightarrow S = \text{constante} \quad (\text{I.29})$$

- Si S prend n modalités α_i , chacune avec une probabilité P_i

$$H(S) \leq \text{Log } n \text{ égalité dans le cas de l'équiprobabilité des modalités}$$

- $\forall S_1, S_2 \in \mathcal{P}(\Sigma)$ la propriété de sous-additivité s'écrit :

$$H(S_1 \cup S_2) \leq H(S_1) + H(S_2) \dots\dots \quad (\text{I.30})$$

avec égalité si S_1 et S_2 sont statistiquement indépendants.

- La propriété de monotonie s'écrit :

a) si $X : \Omega \rightarrow M_X$ et $Y : \Omega \rightarrow M_Y$ sont deux variables définies sur Ω alors :

$$X \leq Y \Rightarrow H(X) \geq H(Y)$$

b) Si $S_1, S_2 \in \mathcal{P}(\Sigma)$ alors :

$$S_1 \subseteq S_2 \Rightarrow H(S_1) \leq H(S_2) \quad (\text{I.32})$$

L'entropie de SHANNON vérifie également, les autres propriétés classiques d'une entropie :

- $H(1) = 0$: une expérience ne présente pas d'incertitude si elle prévoit un seul résultat.

- $H(P_1, \dots, P_n) = H(P_j, \dots, P_{jn})$, j étant une permutation de 1 à n . L'ordre ne doit pas affecter l'incertitude d'un événement (symétrie).
- $H(1/n, \dots, 1/n) < H(1/m, \dots, 1/m)$ si $n < m$: l'incertitude augmente avec le nombre d'événements si ceux-ci sont équiprobables.
- $H(P_1, \dots, P_n, 0) = H(P_1, \dots, P_n)$: une expérience qui prévoit un résultat de probabilité nulle n'augmente pas l'incertitude sur le système (extensibilité).
- $H(P_1, P_2, P_3, \dots, P_n) \geq H(P_1 + P_2, P_3, \dots, P_n)$: une expérience de dimension n a plus d'incertitude que celle de dimension plus faible.

- $\lim_{p \rightarrow 0^+} H(p, 1-p) = 0$: continuité
- $H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$: normalisation

III.2.5 Les propriétés de l'entropie de Max :

Nous allons montrer par la suite que l'utilisation d'une telle entropie peut conduire à la mise en évidence de certaines relations causales entre variables du système qui ne peuvent être prises en compte par l'utilisation de l'entropie de SHANNON. Ses propriétés sont les suivantes :

- Propriété 1 : $H_M(1) = 0$
- Propriété 2 : $H_M(P_1, P_2, \dots, P_n) = H_M(P_j, \dots, P_{jn}) \quad \forall j$ une permutation de 1 à n .
- Propriété 3 : $n \leq m \Rightarrow H_M(1/n, \dots, 1/n) \leq H_M(1/m, \dots, 1/m)$
- Propriété 4 : $H_M(P_1, \dots, P_n) = H_M(P_1, \dots, P_n, 0)$
- Propriété 5 : $H_M(P_1, P_2, P_3, \dots, P_n) \geq H_M(P_1+P_2, P_3, \dots, P_n)$, en effet deux cas peuvent se présenter :
 - . si $P_1+P_2 \leq P_M$ alors : $H_M(P_1, P_2, P_3, \dots, P_n) = H_M(P_1+P_2, P_3, \dots, P_n)$
 - . si $P_1+P_2 > P_M$ alors : $H_M(P_1, P_2, P_3, \dots, P_n) = 1 - P_M > H_M(P_1+P_2, \dots, P_n) = 1 - (P_1+P_2)$
- Propriété 6 : $H_M(X \cup Y) \leq H_M(X) + H_M(Y)$

Preuve :

$$\sum_j \sum_i P_{ij} = 1 = \sum_{i \neq i_0} \sum_{j \neq j_0} P_{ij} + P_{i_0} + P_{j_0} - P_{i_0 j_0} \Rightarrow P_{i_0} + P_{j_0} - P_{i_0 j_0} - 1 \geq 0$$

puisque $\sum_{i \neq i_0} \sum_{j \neq j_0} P_{ij} \geq 0$, ceci est vrai $\forall i_0 \in \{1, \dots, n\}$ et $j_0 \in \{1, \dots, n\}$

et particulièrement pour (i_0, j_0) tels que $P_{i_0} = \max_i P_i$ et $P_{j_0} = \max_j P_j$

$$\text{d'où : } (1 - P_{i_0 j_0}) + (1 - P_{i_0 j_0}) \geq 1 - P_{i_0 j_0} \geq 1 - P_{Mij} = H_M(X \cup Y)$$

$$H_M(X) + H_M(Y) \geq H_M(X \cup Y)$$

L'égalité étant obtenue lorsque :

$$\sum_{i \neq i_0} \sum_{j \neq j_0} P_{ij} = 0 \quad \text{et} \quad P_{i_0 j_0} = P_{Mij}$$

- Propriété 7 : $H_M(X \cup Y) \geq \max [H_M(X), H_M(Y)]$

Preuve :

$$P_{i_0 j_0} = \max_{i,j} P_{ij} \leq \min (P_{i_0}, P_{j_0}) \leq \min (P_{M i}, P_{M j})$$

$$1 - P_{i_0 j_0} \geq 1 - \min (P_{i_0}, P_{j_0}) \geq 1 - \min (P_{M i}, P_{M j}) = \max [H(X), H(Y)]$$

$$H_M(X \cup Y) \geq \max [H_M(X), H_M(Y)]$$

- Propriété 8 :

$$H_M(P, 1-P) = 0 \quad (\text{continuité})$$

$$P \rightarrow 0^+$$

- Propriété 9 :

$$H_M\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}$$

Nous constatons que H_M vérifie la plupart des propriétés classiques. Seule la propriété de normalisation ne se trouve pas vérifiée par une telle entropie. Or ceci ne pose aucun problème particulier étant donné qu'il est toujours possible de normaliser une grandeur en la multipliant par un coefficient approprié.

III.2.6 Entropie d'une variable continue

Une variable x , à un instant t donné, peut prendre toute valeur comprise dans un certain

intervalle (fini ou infini) et obéit à une loi de probabilité de densité $P(X)$. L'extension de la notion d'entropie aux variables continues a été introduite par SHANNON. Sous l'hypothèse de la stationnarité du système, l'entropie d'une variable à un instant donné ne dépend pas du temps. Notons x_t une telle variable obéissant à la loi $P(x)$ telle que :

$$\text{Prob} [x \leq x_t \leq x + dx] = P(x) dx$$

Dans le cas discret, l'entropie s'écrit :

$$\sum_i P_i \log P_i$$

il paraît naturel, avec SHANNON, de passer à la limite, et d'écrire dans le cas des variables continues :

$$H_C(x) = - \int_{-\infty}^{+\infty} P(x) \cdot \text{Log } P(x) dx \quad (\text{I. 33})$$

$$\text{avec } \int_{-\infty}^{+\infty} P(x) dx = 1$$

condition qui exprime le fait que $p(x)$ est une loi de probabilité. Or d'un point de vue mathématique, une telle condition ne suffit à garantir la convergence de l'intégrale précédente. Toutefois, cette convergence est assurée sous certaines conditions concernant $P(x)$ [OSWALD 86].

Quelle interprétation physique peut-on associer à l'expression (I.33). Prenons deux simples exemples : ($P(x)$ successivement un échelon borné et une loi normale centrée).

a)

$$P(x) = \begin{cases} \frac{1}{A}, & \text{pour } 0 \leq x \leq A \\ 0, & \text{pour } x < 0 \text{ et } x > A \end{cases}$$

Le calcul de l'entropie dans ce cas s'écrit :

$$H_C(x) = \text{Log } A$$

Selon les valeurs de A (≥ 1 , < 1), une telle entropie peut être ≥ 0 ou < 0 .

b) une loi normale centrée :

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Le calcul de l'entropie continue donne :

$$H_C(x) = \text{Log}(\sigma\sqrt{2\pi e})$$

Là encore, selon les valeurs de σ ,

$$\left(> \frac{1}{\sqrt{2\pi e}}, \leq \frac{1}{\sqrt{2\pi e}} \right)$$

$H_C(x)$ peut être > 0 ou ≤ 0 . Quelle interprétation physique peut-on donner à une entropie nulle ou négative ? D'autant plus qu'une autre difficulté apparaît lorsque x subit une transformation ($x = g(z)$) [OSWALD 86]. En effet, dans ce cas on a :

$$H'_C(x) = - \int P[g(z)] \text{Log} P[g(z)] dz \quad (\text{I.34})$$

qui diffère de :

$$H_C(x) = - \int P[g(z)] \text{Log} P[g(z)] g'(z) dz$$

OSWALD dit à ce propos "Une relation telle que $H_C(x) = \text{Log} A$ fait dépendre la quantité d'information résultant de la connaissance des états du système de la valeur absolue A de l'amplitude du phénomène. Or, il est clair que le volume d'une émission sonore (au sens des acousticiens) est sans corrélation avec l'information transportée, de la même façon, changer la dimension des lettres d'un texte ne saurait en modifier la teneur ou le sens". il en est de même pour ROUBINE [ROUBINE 1970] qui s'interroge sur l'intérêt d'étendre aux systèmes continus la notion d'entropie. Or SHANNON justifie une telle extension en affirmant que c'est un outil de calcul commode, et dont les difficultés disparaissent lorsque on fait apparaître l'information mutuelle entre variables ou sous-systèmes corrélés. En effet, dans ce cas, l'information mutuelle fait intervenir le calcul de deux entropies dont les valeurs négatives de l'intégrale s'éliminent par soustraction et on obtient alors la propriété de positivité de l'entropie. Enfin, le calcul de l'entropie d'une variable continue par discrétisation est essentiellement pragmatique et n'est pas sans inconvénient et impose des choix judicieux pour que cette discrétisation n'affecte pas les résultats de l'analyse [DUFOUR 79]. Par ailleurs $H(x)$ apparaît, à un coefficient près, comme un estimateur de $H_C(x)$ [YVINEC 78].

III.2.7 Pourquoi s'intéresse-t-on à l'entropie de SHANNON ?

L'entropie de SHANNON a été utilisée avec succès dans le domaine des télécommunications, et en particulier dans les problèmes de codage, et dans bien d'autres domaines tels que l'analyse des données, classification, reconnaissance des formes, questionnaires, ... etc. Or, dans un souci de généralisation de l'entropie de SHANNON, ou bien pour des problèmes bien particuliers, certains auteurs ont proposé d'autres entropies dont les plus connues sont résumées dans la liste ci-dessous.

a) Entropie de RENYI :

$$H_r(P_1, \dots, P_n) = \frac{1}{1-r} \text{Log} \left(\sum_1^n P_i^r \right) \quad (\text{I.35})$$

avec r une constante différente de l'unité et

$$\lim_{r \rightarrow 1} H_r(P_1, \dots, P_n) = H_S(P_1, \dots, P_n) \quad : \text{ entropie de Shannon}$$

H_r est considérée comme une généralisation de l'entropie de SHANNON.

b) Entropie de DAROCZY (ou entropie d'ordre β) :

$$H_\beta(P_1, \dots, P_n) = \frac{1}{2^{1-\beta} - 1} \left(\sum_1^n P_i^\beta - 1 \right) \quad (\text{I.36})$$

• avec $\beta \neq 1$ et

$$\bullet \lim_{\beta \rightarrow 1} H_\beta(P_1, \dots, P_n) = H_S(P_1, \dots, P_n)$$

c) Entropie NORMEE :

$$H_R(P_1, \dots, P_n) = \frac{R}{R-1} \left[1 - \left(\sum_1^n P_i^R \right)^{1/R} \right], \quad \text{avec } R \neq 1 \quad (\text{I.37})$$

d) Entropie hyperbolique :

$$H_h (P_1, \dots, P_n) = \sum_1^n \frac{1}{P_i} - 1 \quad (I.38)$$

e) Entropie de SHARMA-AUTOR :

$$H_f (P_1, \dots, P_n) = \sum_j (P_1 + \dots + P_j) f \left(\frac{P_j}{P_1 + \dots + P_j} \right) \quad (I.39)$$

f) Entropie du type α, β :

$$H_{\alpha, \beta} (P_1, \dots, P_n) = - \frac{1}{2} \text{Log} \left(\frac{\sum_1^n P_i^{\alpha+\beta}}{\sum_1^n P_i^\beta} \right) \quad (I.40)$$

On constate que dans l'ensemble de ces entropies, on fait apparaître à chaque fois un paramètre dont le choix reste toujours arbitraire. Dans tous les formalismes mathématiques conduisant à ces entropies, aucune signification physique n'a été donnée pour de tels paramètres si bien que l'utilisateur se trouve confronté au choix de l'entropie la mieux adaptée à son analyse. En effet, comme le dit G. JUMARIE [1978] à propos de l'entropie de RENYI "Il n'y a absolument rien dans le formalisme mathématique conduisant à H_r , qui permette de préciser d'une façon ou d'une autre la signification concrète de r si bien que le technicien ne dispose d'aucun critère systématique de choix lui permettant de préférer H_r à H_s , ou vice versa, dans un problème réel de transmission de message".

Le même problème se pose pour l'emploi de telles entropies dans l'analyse structurale. En effet, celles-ci ne vérifient pas toujours la propriété de la sous-additivité et ne peuvent fournir des indices de couplage, ... etc. facilement interprétables. C'est pourquoi, nous utiliserons désormais l'entropie de SHANNON, puisque celle-ci ne fait intervenir aucun paramètre particulier, vérifie la propriété de la sous-additivité et conduit à des indices dont l'interprétation se trouve aisée. Nous l'utiliserons quelquefois conjointement avec l'entropie du Max, pour analyser les couplages entre variables ou sous-systèmes du système global.

IV - APPLICATION DES OUTILS DE L'ANALYSE DES DONNÉES ET DE LA THÉORIE DE L'INFORMATION A L'AUTOMATIQUE

IV.1 L'apport des outils de l'analyse des données

Nous ne ferons pas ici une description détaillée de telles méthodes pour une présentation approfondie, on pourra se reporter aux ouvrages spécialisés [BENZECRI 80] [SAPORTA 78] [L. LEBART 82]. Lorsque, on est amené à analyser un système à travers ses données, deux objectifs principaux peuvent être visés :

- recherche de structures : aspect décomposition ;
- recherche de relations fonctionnelles : aspect modélisation.

Les méthodes que nous avons présentées au paragraphe III-1 répondent en partie à ces deux objectifs. En effet, dans le premiers cas, on fait appel aux méthodes factorielles (analyse en composantes principales et l'analyse des correspondances), dans le second cas, les méthodes de régression et d'analyse canonique apportent une réponse à cette nécessité.

IV.1.1 L'apport des méthodes factorielles

a) Analyse en composantes principales : On dispose d'un ensemble d'observations recueillies sur le système. La masse importante des informations à traiter constitue un obstacle pour leurs exploitations. Pour dégager les caractéristiques importantes des informations que recèlent ces données, une projection de l'ensemble des points représentatifs du système dans un espace de dimension réduite est alors opérée. La meilleure représentation est celle qui conduit au nuage projection le plus proche du nuage initial. La qualité de la projection est mesurée par le rapport des inerties des deux nuages. Cette méthode permet donc :

- de mettre en évidence les proximités entre variables qui peuvent se traduire en termes de dépendance entre celles-ci ;
- de faire apparaître les relations suivant la hiérarchie de leur importance [DUFOR 79] ;
- dans le cadre d'une modélisation, de donner une vision synthétique de l'information obtenue sur le modèle et de faciliter le dialogue et la critique des modèles complexes élaborés par des équipes pluridisciplinaires [GENTIL 84] ;
- une interprétation facile d'emploi pour l'organisation, la structuration des données tout en soulignant leur ressemblance ou dissemblance.

Enfin, une telle méthode peut être utilisée conjointement avec d'autres approches pour faciliter les aspects décomposition et modélisation des systèmes complexes.

b) L'analyse factorielle des correspondances : Cette méthode se distingue de la précédente par la nature des tableaux qu'elle traite (tableau de contingence) et utilise la métrique du χ^2 au lieu de la distance euclidienne. Elle tient compte du caractère probabiliste des données et la symétrie existant entre les lignes et les colonnes. Le choix de la distance du χ^2 est lié au principe d'équivalence distributionnelle. Le regroupement de deux lignes de même profil s'opère sans aucune perte d'information. La représentation du nuage de points associé au système, sur les axes factoriels les plus significatifs conduit à des résultats interprétables lorsque ceux-ci possèdent une structure facilement identifiable. Même si certaines interprétations paraissent "évidentes" au premier coup d'œil, il est indispensable d'utiliser les "aides à l'interprétation" qui doivent être éditées conjointement avec les résultats obtenus [VOLLE 76] et permettent de tirer le plus grand profit possible de l'analyse. Les aides à l'interprétation peuvent être de natures diverses :

- liste des valeurs propres, le pourcentage de la variance totale expliquée par chaque axe ainsi que le cumul de ces pourcentages ;
- un tableau faisant apparaître pour chaque variable sa contribution à l'inertie totale, sa contribution à l'inertie expliquée par le premier axe ;
- la qualité de la représentation, ... etc.

Par ailleurs, si la forme du nuage projeté sur le plan factoriel peut être décomposé en "paquets", cela signifie que le tableau de contingence peut être scindé en autant de correspondances que de paquets. Ceci conduit naturellement à analyser séparément des sous-tableaux obtenus par décomposition du tableau de contingence initial. Cette façon de procéder rejoint la notion de décomposition des systèmes en sous-systèmes faibles couplés et autorise l'application des méthodes de modélisation par modèles partiels, commande décentralisée, commande hiérarchisée, ... etc.

IV.1.2 L'apport des méthodes de régression :

Appliquées à notre domaine, celles-ci peuvent étudier les relations d'entrée-sortie d'un système linéaire. Autrement dit, ils peuvent nous renseigner sur l'intensité des liaisons linéaires entre variables et groupe de variables du système. Par ailleurs, de nombreux modèles non-linéaires se ramènent facilement au modèle linéaire par des transformations simples [SAPORTA 78]. Ainsi par exemple, le modèle entrée-sortie $Y \# \alpha X^\beta$ utilisé en économétrie (élasticité constante de Y par rapport à X, β coefficient d'élasticité) peut se ramener à un modèle linéaire, en posant $U = \text{Log } Y$ et $V = \text{Log } X$, et s'écrit : $U \# \text{Log } \alpha + \beta V$. Il en est de même pour un modèle logistique souvent utilisé pour rendre compte des variations d'un taux de réponse Y (compris entre 0 et 1) en fonction d'une "excitation" X :

$$Y \# \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}} \quad \text{se linéarise en posant}$$

$$U = \text{Log} \frac{Y}{1-Y} \quad \text{et devient} \quad U \# \alpha + \beta X$$

Enfin, nous retenons que cette méthode utilise un modèle linéaire (ou linéarisable) pour décrire certains aspects de dépendance entre variables aléatoires (ou entre variables aléatoires et variables certaines) et pour tester certaines hypothèses concernant cette dépendance.

IV.1.3 L'apport de l'analyse canonique

Cette approche constitue une généralisation de l'approche précédente, dans la mesure où l'on tente "d'expliquer" un groupe de variables au moyen d'un autre groupe de variables. Elle cherche par conséquent, à synthétiser les inter-relations existant entre deux sous-ensembles de variables en cherchant les combinaisons linéaires des variables du premier sous-ensemble les plus corrélées à des combinaisons linéaires des variables du second sous-ensemble [LEBART 75]. Cette méthode peut être utilisée pour analyser également les relations entrée-sortie d'un système complexe et constitue une aide précieuse pour dégager un ensemble minimal de variables caractérisant le comportement de ce système. Par ailleurs, l'application de cette méthode à différents jeux de variables permettrait de suivre la structure d'un système complexe au cours du temps en analysant les données sur des horizons successifs [DUFOR 79] [STAROSWIECKI 79, 81].

La plupart des méthodes précédemment décrites se basent sur l'hypothèse de linéarité des relations liant les variables du système. Chacune d'elles est adaptée à un certain type de variables (quantitatives, qualitatives, ... etc.) et possède sa propre spécificité. Nous sommes amenés parfois à pratiquer différentes analyses, du même tableau, pour ressortir les résultats escomptés. C'est pourquoi, nous allons utiliser les outils de la théorie de l'information qui se prêtent bien à l'analyse des systèmes.

IV.2 L'apport des outils de la théorie de l'information

Nous allons faire une synthèse des principaux résultats de la théorie de l'information (T.I.) applicables à l'analyse structurale des systèmes. Nous montrerons ensuite que cette approche possède certains avantages, que nous essaierons de mettre en valeur.

IV.2.1 Les transinformations

La propriété de sous-additivité de l'entropie formulée par :

$$\forall S_1, S_2 \in \mathcal{P}(\Sigma) : H(S_1 \cup S_2) \leq H(S_1) + H(S_2)$$

est interprétée par le fait que dans $H(S_1) + H(S_2)$ apparaît deux fois la même quantité d'information se trouvant aussi bien dans S_1 que dans S_2 et supposée comme ayant été transmise. Pour mesurer une telle transmission (ou échange) d'information, on définit de façon classique à partir de l'entropie :

a) La Transinformation Interne :

$$\begin{aligned} \dot{I} : \mathcal{P}(\Sigma) &\longrightarrow \mathbb{R} \\ S &\longrightarrow \dot{I}(S) = \sum_{X_i \in S} H(X_i) - H(S) \end{aligned} \quad (I.41)$$

relation qui s'interprète également comme la quantité d'information échangée à l'intérieur d'un sous-système constitué d'un groupe de variables du système global.

b) La Transinformation Externe définie par :

$$\begin{aligned} \leftrightarrow I : \mathcal{P}(\mathcal{P}(\Sigma)) &\longrightarrow \mathbb{R} \\ R = \{R_1, R_2, \dots, R_k\} &\longrightarrow \leftrightarrow I(R) = \sum_{i=1}^k H(R_i) - H\left(\bigcup_{i=1}^k R_i\right) \end{aligned} \quad (I.42)$$

relation qui mesure la quantité d'information échangée ou transmise entre plusieurs sous-systèmes (ou groupes de variables).

Les relations, entre transinformations internes, externes et l'entropie vérifient certaines propriétés intéressantes que l'on donne ci-dessous.

IV.2.2 Les propriétés des transinformations

Celles-ci sont, en grande partie formulées par ASHBY [1965-69] dans le cas où H est une entropie et généralisées pour d'autres fonctions (semi-valuation monotone ou non, semi-modulation positive monotone ou non) par Z. ABID [1979] et V. TORO [1982].

Propriété 1 :

$$- \forall R = \{P_1, \dots, P_k\} \in \mathcal{P}(\mathcal{P}(\Sigma))$$

$$\vec{I}(R_1 : R_{2i} \dots R_k) \geq 0 \quad (\text{I.43})$$

$$- \forall R_i \in \mathcal{P}(\Sigma) : \dot{I}(R_i) \geq 0 \quad (\text{I.44})$$

Cette propriété indique que la transinformation aussi bien interne qu'externe est toujours positive. Elle ne s'annule que dans le cas où les variables (ou groupes de variables) sont indépendantes entre elles. Ceci constitue un avantage non négligeable lorsqu'il s'agit d'analyser les relations entre variables ou groupes de variables pour une décomposition ou modélisation du système.

Propriété 2 :

$$\forall R, R' \in \mathcal{P}(\mathcal{P}(\Sigma)), R' \subseteq R$$

$$\bullet \overset{\leftrightarrow}{I}(R') \leq \overset{\leftrightarrow}{I}(R) \quad (\text{I.45})$$

$$\bullet \overset{\leftrightarrow}{I}(R) = 0 \Rightarrow \overset{\leftrightarrow}{I}(R_i : R_j) = 0 \quad \forall R_i, R_j \in R \quad (\text{I.46})$$

La première affirmation signifie que l'information totale véhiculée par un système ne peut que diminuer si l'on supprime certaines de ses composantes.

La deuxième affirmation signifie que si l'information échangée globalement entre les différents sous-systèmes est nulle, alors il n'y aura aucun échange d'information entre ces sous-systèmes pris deux à deux. Par ailleurs, on peut remarquer que la réciproque n'est pas vraie. C'est pourquoi, l'analyse des couplages deux à deux s'avère parfois inopérante dans la mise en évidence des relations structurales.

Propriété 3 :

$$\forall S, S' \in \mathcal{P}(\Sigma), S' \subseteq S, \text{ on a :}$$

$$\bullet \text{ si } \dot{I}(S') \leq \dot{I}(S) \quad (\text{I.47})$$

$$\bullet \text{ si } \dot{I}(S) = 0 \Rightarrow \overset{\leftrightarrow}{I}(X_i : X_j) = \dot{I}(\{X_i, X_j\}) = 0 \quad \forall X_i, X_j \in S \quad (\text{I.48})$$

L'interprétation d'une telle propriété est similaire à la précédente. En effet, dans ce cas, on s'intéresse à l'échange d'information à l'intérieur d'un sous-système (niveau local) contrai-

rement au cas précédent où l'on met en évidence l'échange d'information entre sous-systèmes (niveau global). On constate également que la réciproque de la deuxième affirmation n'est pas vraie.

Propriété 4 : Soient $S_1, S_2, \dots, S_k \in \mathcal{P}(\Sigma)$ on a :

$$\begin{aligned}
 \bullet \quad \overset{\leftrightarrow}{I}(S_1 : S_2 : \dots : S_k) &= \overset{\leftrightarrow}{I}(S_1 : S_2 : \dots : S_r) + \overset{\leftrightarrow}{I}\left(\bigcup_1^r S_i : \bigcup_{r+1}^k S_i\right) \\
 &\quad + \overset{\leftrightarrow}{I}(S_{r+1} : \dots : S_k), \quad \forall r=2, \dots, k-2 \\
 &= \overset{\leftrightarrow}{I}(S_1 : S_2) + \overset{\leftrightarrow}{I}(S_1 \cup S_2 : S_3) + \overset{\leftrightarrow}{I}(S_1 \cup S_2 \cup S_3 : S_4) + \dots \\
 &\quad \dots + \overset{\leftrightarrow}{I}(S_1 \cup S_2 \cup \dots \cup S_{k-1} : S_k)
 \end{aligned} \tag{I.49}$$

Cette propriété met en évidence la possibilité d'une décomposition de l'information transmise entre sous-systèmes.

Propriété 5 : \forall la partition $S = \{S_1, S_2, \dots, S_k\} \in \mathcal{IP}(\Sigma)$ on a :

$$\overset{\leftrightarrow}{I}(S_1 : S_2 : \dots : S_k) + \sum_{i=1}^k \overset{\bullet}{I}(S_i) = \text{cte} = \overset{\leftrightarrow}{I}(X_1 : X_2 : \dots : X_N) \tag{I.50}$$

L'information totale transmise dans un système peut être décomposée en information transmise au sein des sous-systèmes (transinformation interne) et information transmise entre sous-systèmes (transinformation externe). Cette relation est analogue à celle donnée par le théorème de König-Huygens relatif à la décomposition des moments d'inertie, utilisée en mécanique classique et en Analyse des données.

Propriété 6 : Soient $R = \{R_1, \dots, R_r\}$ et $S = \{S_1, \dots, S_k\} \in \mathcal{IP}(\Sigma)$ on a :

$$R \leq S \Rightarrow \begin{cases} \overset{\leftrightarrow}{I}(R) \geq \overset{\leftrightarrow}{I}(S) \\ \sum_1^r \overset{\bullet}{I}(R_i) \leq \sum_1^k \overset{\bullet}{I}(S_i) \end{cases} \tag{I.51}$$

$$\sum_1^r \overset{\bullet}{I}(R_i) \leq \sum_1^k \overset{\bullet}{I}(S_i) \tag{I.52}$$

Plus la décomposition du système est fine, plus la transinformation entre les sous-systèmes augmente et par conséquent la transformation à l'intérieur des sous-systèmes diminue.

Propriété 7 :

$$\forall R = \{R_1, \dots, R_i, \dots, R_j, \dots, R_k\} \in \mathcal{P}(\Sigma)$$

et soit $R' = \{R_1, \dots, R_i \cup R_j, \dots, R_k\} \in \mathcal{P}(\Sigma)$ on a :

$$I(R') = I(R) - I(R_i : R_j) \quad (I.53)$$

Le regroupement de deux sous-systèmes diminue la transinformation externe d'une quantité égale à celle qui existe initialement avant regroupement.

Propriété 8 : Soit $R' = \{R_1, R_2, \dots, R_i \setminus S, S, \dots, R_k\}$, avec $S \subseteq R_i$, $S \neq R_i$, $S \neq \emptyset$, $\text{card}(R_i) \geq 2$ on a :

$$\overset{\leftrightarrow}{I}(R') = \overset{\leftrightarrow}{I}(R) + \overset{\leftrightarrow}{I}(R_i \setminus S : S) \quad (I.54)$$

Lorsqu'on décompose un sous-système en deux sous-systèmes, l'information totale entre sous-systèmes initiaux augmente de la même quantité que celle générée par l'opération de décomposition.

IV.2.3 Entropie et transinformation conditionnelle

IV.2.3.1 Définitions

Soient $X, Y \in \mathcal{P}(\Sigma)$, on a vu que : $I(X : Y) = H(X) + H(Y) - H(X \cup Y)$ mesure la quantité d'information échangée (ou transmise) entre X et Y . La question qui se pose est celle de savoir comment évaluer l'information qui reste à connaître sur Y une fois que X est connu. La réponse à cette question est simple : l'information restant dans Y , après la connaissance de X est donnée par :

$$H(Y) - I(X : Y) = H(Y) - H(X) - H(Y) + H(X \cup Y) = H(X \cup Y) - H(X)$$

ceci nous conduit à définir l'entropie X conditionnelle par :

$$\forall Y \in \mathcal{P}(\Sigma) :$$

$$\begin{aligned} H(. / X) : \mathcal{P}(\Sigma) &\longrightarrow \mathbb{R} \\ Y &\longrightarrow H(Y / X) = H(X \cup Y) - H(X) \end{aligned} \quad (I.55)$$

En termes d'information, cette relation interprète comme la quantité moyenne d'infor-

mation que fournit encore Y lorsqu'on rend constante la partie du système concernée par X [ASHBY 65]. C'est aussi la quantité d'information qui reste à connaître sur l'ensemble des variables Y, une fois que l'on possède déjà l'information fournie par X. Par ailleurs, nous pouvons écrire une expression intéressante de $H(Y/X)$ qui nous servira pour procéder à des analyses locales. En effet, si on note $M_X = \{\alpha_i, i=1, \dots, n\}$ et $M_Y = \{\beta_j, j=1, \dots, m\}$ les ensembles de modalités de X et de Y avec

$P_i = P(X = \alpha_i)$, $P_{.j} = P(Y = \beta_j)$, $P_{ij} = P(X = \alpha_i, Y = \beta_j)$ et $P_{i/j} = P(Y = \beta_j/X = \alpha_i)$ alors $H(Y/X)$ peut s'écrire :

$$\begin{aligned}
 H(Y/X) &= H(X \cup Y) - H(X) \\
 &= - \sum_i \sum_j P_{ij} \text{Log } P_{ij} - \sum_i P_i \text{Log } P_i = \sum_i \sum_j P_{ij} \text{Log } P_{j/i} \\
 &= \sum_i P_i \left[\sum_j P_{j/i} \text{Log } P_{j/i} \right] = \sum_i P_i \cdot H(Y/X = \alpha_i) \\
 &= \sum_{i=1}^n P(X=\alpha_i) H(Y/X = \alpha_i)
 \end{aligned} \tag{I.56}$$

Ainsi, $H(Y/X)$ l'entropie globale de Y sachant X est la somme de n entropies locales $H(Y/X = \alpha_i)$; $i = 1, \dots, n$; pondérées par leurs poids respectifs. C'est la quantité moyenne d'information restant à évaluer sur l'ensemble des variables Y, lorsque l'information fournie par l'ensemble des variables X est connue. Ainsi par exemple, dans une machine à courant continu où sont mises en jeu des variables comme la tension d'induit U, le courant d'induit I, le courant inducteur i, vitesse de rotation N, ..., $H(U \cdot I / i \cdot N)$ serait la quantité moyenne d'information que l'on peut encore extraire de U et de I lorsque le courant d'excitation et la vitesse d'entraînement sont constantes. Cette possibilité de "garder" constantes ou de "fixer" certaines variables du système peut être mise à profit aux moyens des transinformatives X conditionnelles que l'on peut définir par :

$$\begin{aligned}
 \dot{I}_X &: \mathcal{P}(\Sigma) \longrightarrow \mathbb{R} \\
 Y &\longrightarrow \dot{I}_X(Y) = \sum_{A \in Y} H(A/X) - H(Y/X)
 \end{aligned} \tag{I.57}$$

$$\begin{aligned}
 \leftrightarrow I_X &: \mathcal{P}(\mathcal{P}(\Sigma)) \longrightarrow \mathbb{R} \\
 R = \{R_1, \dots, R_k\} &\longrightarrow \leftrightarrow I_X(R) = \sum_{i=1}^k H(R_i/X) - H\left(\bigcup_{i=1}^k R_i/X\right)
 \end{aligned} \tag{I.58}$$

Ces deux quantités mesurent les transinformations lorsque l'ensemble des variables est fixé. Ceci débouche sur la définition de la transinformation directe [ASHBY 65] qui s'écrit dans le cas où l'on conditionne celle-ci entre deux variables avec le reste du système :

$$\begin{aligned} \overset{\leftrightarrow}{I}_D(X_i : X_j) &= \overset{\leftrightarrow}{I}_{\Sigma \setminus \{X_i, X_j\}}(X_i : X_j) \\ &= H(X_i / \Sigma \setminus \{X_i, X_j\}) - H(X_i / \Sigma \setminus X_i) \\ &= H(\Sigma \setminus X_i) + H(\Sigma \setminus X_j) - H(\Sigma \setminus \{X_i, X_j\}) - H(\Sigma) \end{aligned} \quad (I.59)$$

En termes d'information, cette quantité mesure la transmission d'information directe entre X_i et X_j sans passer par les autres variables du système. Cette expression peut être naturellement généralisée pour le cas multi-dimensionnel, c'est-à-dire :

$$\forall X, Y \in \mathcal{P}(\Sigma) \text{ on définit :} \quad (I.60)$$

$$\overset{\leftrightarrow}{I}_D(Y : X) = \overset{\leftrightarrow}{I}_{\Sigma \setminus \{X, Y\}}(Y : X) = H(\Sigma \setminus X) + H(\Sigma \setminus Y) - H(\Sigma \setminus \{X, Y\}) - H(\Sigma)$$

relation qui s'interprète également comme la quantité d'information qui transite directement entre les variables vectorielles X et Y sans passer par l'intermédiaire des variables constituant le reste du système. Une telle relation peut être utilisée en modélisation ou en décomposition des systèmes et possède certains avantages et certains inconvénients que nous préciserons ultérieurement.

IV.2.3.2 Quelques propriétés de l'entropie conditionnelle et des transinformations

Nous énumérons quelques propriétés classiques de ces fonctions :

Propriété 1 : $\forall X, Y \in \mathcal{P}(\Sigma)$:

$$0 \leq H(Y/X) \leq H(Y) \quad (I.61)$$

En effet, la connaissance d'un groupe de variables du système ne peut que diminuer l'incertitude sur n'importe quel autre groupe de variables du système.

Pour les deux cas limites, on a :

- $H(Y/X) = H(Y)$ indépendance totale entre Y et X .

- $H(Y/X) = 0$: signifie que la quantité d'information qui reste à connaître est nulle.

Propriété 2 : $\forall X, Y, Z \in \mathcal{P}(\Sigma), Y \subseteq X$, on a :

$$H(Y/Z) \leq H(X/Z) \quad (I.62)$$

Cette propriété exprime le fait que H étant une fonction monotone par rapport à l'inclusion $H(. / Z)$ l'est également.

Propriété 3 : $\forall X, Y, Z \in \mathcal{P}(\Sigma) : Z \subseteq X$ on a :

$$H(Y/Z) \geq H(Y/X) \quad (I.63)$$

L'incertitude sur un groupe de variables diminue lorsque la connaissance sur les autres variables du système augmente.

Propriété 4 : $\forall X, X', Y, Y' \in \mathcal{P}(\Sigma) : X' \subseteq X, Y' \subseteq Y$ on a :

$$\begin{matrix} \leftrightarrow & & \leftrightarrow \\ I(X' : Y') & \leq & I(X : Y) \end{matrix} \quad (I.64)$$

Cette propriété signifie que plus la dimension des sous-systèmes augmente, plus l'échange d'information entre eux devient important.

IV.2.4 Le principe de conditionnement uniforme [GARNER 56, ASHBY 65]

Ce principe appelé quelques fois sous indexation uniforme utilise les concepts de l'entropie et de l'information conditionnelles. Un tel principe peut être énoncé comme suit : "Si l'on conditionne avec le même ensemble de variables, tous les termes d'une identité valable pour H = entropie, qui contient $H(\cdot)$, $H(\cdot / \cdot)$, I, I_α , I, I_β on obtient à nouveau une identité valable pour l'entropie (le double conditionnement étant équivalent au conditionnement avec l'union)". En effet, toutes les affirmations précédentes valables pour l'entropie peuvent être obtenues après conditionnement. Ainsi par exemple : $\forall X, Y, Z \in \mathcal{P}(\Sigma)$

$$i) H(X \cup Y) \leq H(X) + H(Y) \quad : \quad H(X \cup Y / Z) \geq H(X/Z) + H(Y/Z) \quad (I.65)$$

$$ii) I(Y : X) = H(Y) - H(Y/X) \quad : \quad I_Z(Y : X) = H(Y/Z) - H(Y/X \cup Z) \quad (I.66)$$

L'utilisation du principe de conditionnement uniforme doit être faite avec prudence, ainsi par exemple :

$$H(Y \cup X) = H(X) + H(Y) \Rightarrow H(Y \cup X/Z) = H(X/Z) + H(Y/Z) \quad (I.67)$$

peut ne pas être vérifiée.

En effet X et Y sont statistiquement indépendants, l'entropie de leur union conditionnée par Z est inférieure ou égale à la somme des entropies de chacune d'elles conditionnées par le même sous-ensemble Z. Ceci provient du fait que (I.67) n'est pas une identité.

$H(X \cup Y) = H(X) + H(Y) \Rightarrow H(X \cup Y/Z) \leq H(X/Z) + H(Y/Z)$. En effet, on peut bien avoir : $I(X : Y) = 0$ et $I_Z(X : Y) \geq 0$

Pour bien illustrer ces éventualités, prenons un simple exemple :
Soit le tableau de données ci-dessous (fig. I.2)

$$\Sigma = \{X_i, i = 1, \dots, 7\}$$

$$\begin{aligned} \text{Si on pose : } X &= \{X_1, X_3, X_4\} \\ Y &= \{X_2, X_5, X_6\} \\ Z &= \{X_7\} \end{aligned}$$

Le calcul des différentes entropies donne :

$$H(X) = 2.32, H(Y) = 1, H(X \cup Y) = 3.32, H(Z) = 1$$

$$H(X/Z) = 1.52, H(Y/Z) = 0.97, H(X \cup Y/Z) = 2.32 \quad \text{d'où}$$

$$H(X \cup Y) = 3.32 = H(X) + H(Y) \text{ mais } H(X \cup Y/Z) = 2.32 < [H(X/Z) + H(Y/Z)] = 2.49$$

$$\text{de même : } I(X : Y) = 0 \text{ alors que } I_Z(X : Y) = 0.17$$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
ω_0	1	0	1	0	1	0	0
ω_1	1	1	1	0	0	1	0
ω_2	1	0	0	1	1	0	1
ω_3	1	1	0	1	0	1	1
ω_4	1	0	1	1	1	0	0
ω_5	1	1	1	1	0	1	0
ω_6	0	0	1	1	1	0	1
ω_7	0	1	1	1	0	1	1
ω_8	0	0	0	0	1	0	0
ω_9	0	1	0	0	0	1	1

Fig. I.2

IV.2.5 Interprétation graphique d'un système :

Il est possible d'établir une analogie entre l'entropie et la théorie de la mesure qui permet sous quelques réserves une interprétation graphique [MILLER 63, TORO 82]. Comme l'indique également E. ROUBINE [ROUBINE 70] : "les entropies sont des grandeurs positives attachées à des ensembles, donc des mesures de ces ensembles". Cette analogie permet de représenter chaque variable, dans le cas d'une représentation plane, par une surface proportionnelle à son entropie. Comme le signale plus récemment J. OSWALD [OSWALD 86], si on note

$\mu(X), \mu(Y), \mu(X \cap Y), \dots$ les mesures des ensembles $X, Y, X \cup Y, \dots$ on peut alors écrire :

$$\begin{aligned}
 & - H(X) = \mu(X) & - H(Y/X) &= \mu(Y) - \mu(X \cap Y) \\
 & - H(Y) = \mu(Y) & - I(X : Y) &= \mu(X \cap Y) \\
 & - H(X \cup Y) = \mu(X \cup Y) & &
 \end{aligned}
 \tag{I.68}$$

On peut illustrer cette analogie à l'aide de l'exemple ci-dessous, dans lequel chaque variable est représentée par un cercle de surface proportionnelle à son entropie. La surface de l'intersection entre deux cercles représente l'information transmise entre les deux variables mises en jeu.

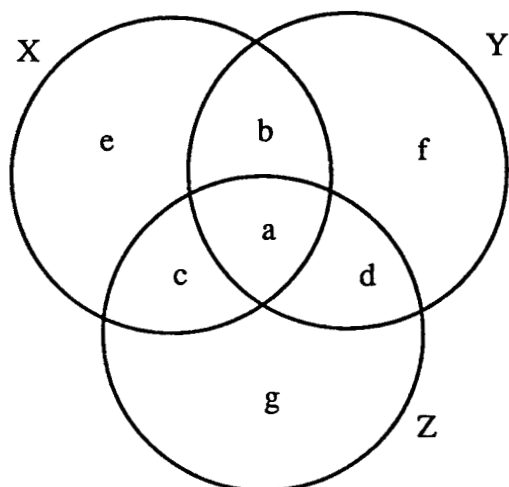


Fig. I.3 : représentation graphique d'un système à 3 variables X, Y et Z .

D'après le parallèle qui existe donc entre l'entropie et la théorie de la mesure, on peut alors écrire que :

$$\begin{aligned}
 & - H(X) = a + b + c + e \\
 & - H(X \cup Y) = a + b + c + d + e + f \\
 & - H(X \cup Y \cup Z) = a + b + c + d + e + f + g \\
 \\
 & - H(X/Y) = e + c \\
 & - H(X/Y \cup Z) = e \\
 & - H(X \cup Y/Z) = b + e + f \\
 \\
 & - I(X : Y) = b + a \\
 & - I(X : Y \cup Z) = a + b + c \\
 & - I_Z(X : Y) = b \\
 & - I(X : Y : Z) = 2a + b + c + d
 \end{aligned}$$

Dans ce dernier cas, la surface a est comptée deux fois à cause de la présence de trois variables. Pour un système à N variables, celle-ci doit être comptée $(N-1)$ fois et correspond à

leur intersection ; de même l'intersection de k variables, lorsque celle-ci existe, doit être comptée (k-1) fois, $\forall k \in \{2, \dots, N\}$. Par ailleurs, dans la mesure où l'analogie entre la théorie de la mesure et l'entropie n'est pas totale, on peut se demander effectivement si à tout système réel, on peut faire correspondre une représentation graphique et vice versa ? La réponse à cette question a été donnée en partie par TORO [1982]. En effet, celui-ci montre que :

- à toute représentation graphique correspond un système réel ;
- à tout système n'a pas de représentation graphique.

En fait, ceci est dû essentiellement à la non-réciprocité de la relation (I.46). En effet, nous allons montrer que si la réciprocité d'une telle relation n'est pas vérifiée, pour un système donné, par l'utilisation de l'entropie de SHANNON, elle peut l'être par contre par l'utilisation de l'entropie de Max. La non réciprocité de la relation (I.46) est caractérisée par :

$\forall i, j \in \{1, \dots, N\} ; i \neq j : I(X_i : X_j) = 0$ mais $I(X_1 : X_2 : \dots : X_N) > 0$ et ceci quelle que soit l'entropie utilisée. En notant $I(X : Y)_S$ et $I(X : Y)_M$ les transinformations calculées respectivement à l'aide de l'entropie de SHANNON et de l'entropie du Max, on remarque que :

$$\text{- si } I(X : Y)_S = 0 \text{ alors } I(X : Y)_M > 0 \quad (1)$$

(I.69)

$$\text{- si } I(X : Y)_M = 0 \text{ alors } I(X : Y)_S > 0 \quad (2)$$

Preuve :

$$\text{Notons } P_{i_0} = \max_i P_i, \quad P_{j_0} = \max_j P_j \quad \text{et} \quad P_{i_0 j_0} = \max_{i, j} P_{i, j}$$

$$\begin{aligned} \text{- } I_S(X : Y) = 0 &\Rightarrow H_S(X) + H_S(Y) = H_S(X \cup Y) \Rightarrow P_{ij} = P_i \cdot P_j, \forall i, j \Rightarrow \\ &P_{i_0 j_0} = P_{i_0} \cdot P_{j_0} \text{ d'où } I_M(X : Y) = H_M(X) + H_M(Y) - H_M(X \cup Y) \\ &= (1 - P_{i_0}) + (1 - P_{j_0}) - (1 - P_{i_0 j_0}) = 1 + P_{i_0 j_0} - P_{i_0} - P_{j_0} = 1 + P_{i_0 j_0} - P_{i_0} P_{j_0} \\ &= (1 - P_{i_0})(1 - P_{j_0}) = H_M(X) \cdot H_M(Y) > 0 \text{ à condition que } H_M(X) \text{ et } H_M(Y) > 0 \end{aligned}$$

d'où (1)

$$I(X : Y)_M = 0 \Rightarrow \sum_{i \neq i_0} \sum_{j \neq j_0} P_{ij} = 0 \quad \text{et} \quad P_{i_0 j_0} + 1 = P_{i_0} + P_{j_0}$$

$$\begin{aligned} I(X, Y)_S &= \left[-P_{i_0} \text{Log } P_{i_0} - \sum_{i \neq i_0} P_i \text{Log } P_i \right] + \left[-P_{j_0} \text{Log } P_{j_0} - \sum_{j \neq j_0} P_j \text{Log } P_j \right] \\ &+ P_{i_0 j_0} \text{Log } P_{i_0 j_0} + \sum_{j \neq j_0} P_{i_0 j} \text{Log } P_{i_0 j} + \sum_{i \neq i_0} P_{ij_0} \text{Log } P_{ij_0} \\ &= -P_{i_0} \text{Log } P_{i_0} - P_{j_0} \text{Log } P_{j_0} + P_{i_0 j_0} \text{Log } P_{i_0 j_0} > 0 \end{aligned}$$

d'où (2)

Interprétation : Si un système n'admet pas une représentation graphique pour l'entropie de SHANNON, cela n'exclut pas l'existence d'une telle représentation pour l'entropie de Max et réciproquement. Nous allons présenter ci-dessous un système de quatre variables binaires, pour lequel il n'y a pas de représentation graphique, dans le cas où on fait appel à l'entropie de SHANNON et tel qu'il est possible de trouver une représentation pour l'entropie de Max.

Exemple : Soit le système combinatoire, défini par le tableau de données (fig. I.4), et correspondant à la fonction logique :

$$X_4 = \bar{X}_1 \bar{X}_2 \bar{X}_3 + \bar{X}_1 X_2 X_3 + X_1 \bar{X}_2 X_3 + X_1 X_2 \bar{X}_3$$

- L'utilisation de l'entropie de SHANNON conduit à :

- . $H(X_i) = 1 \quad \forall i = 1, \dots, 4$
- . $H(X_i X_j) = 2 \quad \forall i \neq j$
- . $H(X_i X_j X_k) = 3 \quad \forall i \neq j \neq k$
- . $H(X_1 X_2 X_3 X_4) = 3$
- . $I(X_i : X_j) = 0 \quad \forall i \neq j$
- . $I(X_i : X_j : X_k) = 0 \quad \forall i \neq j \neq k$
- . $I(X_1 : X_2 : X_3 : X_4) = 1$

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

Fig. I.4

L'examen de ces résultats montre l'impossibilité d'effectuer une représentation graphique puisque il faudrait à la fois quatre cercles isolés, chacun d'une surface égale à 1, qui ne se coupent pas et dont la surface totale doit être égale à 1. Par ailleurs, d'un point de vue physique, cela veut dire que l'état d'une variable est complètement déterminée par la connaissance des états des trois autres variables.

L'utilisation de l'entropie de Max donne :

- . $H(X_i) = 1/2 \quad \forall i = 1, \dots, 4$
- . $H(X_i X_j) = 3/4 \quad \forall i \neq j$
- . $H(X_i X_j X_k) = 7/8 \quad \forall i \neq j \neq k$
- . $H(X_1 X_2 X_3 X_4) = 7/8$
- . $I(X_i : X_j) = 1/4, \quad \forall i \neq j$
- . $I(X_i : X_j : X_k) = 5/8, \quad \forall i \neq j \neq k$
- . $I(X_1 : X_2 : X_3 : X_4) = 9/8$

L'analyse de ces résultats prouve qu'il existe un lien entre les variables prises deux à deux et qu'il est possible de trouver une représentation graphique pour un tel système. En effet, à la figure I.5, on donne une telle représentation géométrique dans laquelle la zone hachurée représente la variable X_4 .

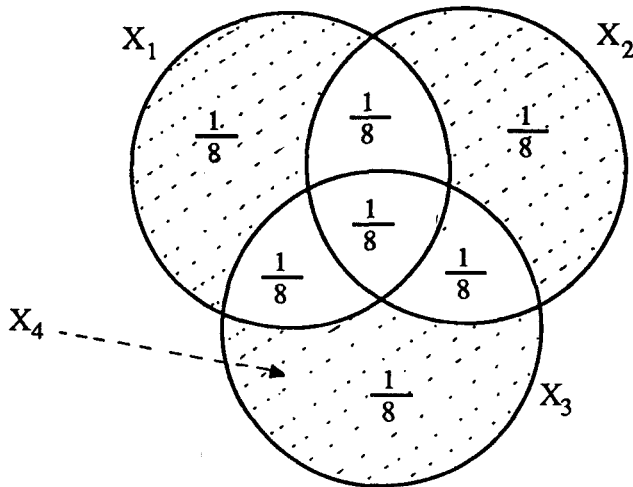


fig. I.5 : Représentation graphique de Σ

En fait, le système étudié est un compteur 3 bits + 1 bit de parité. Ainsi, ce qui semblait initialement un système à quatre variables indépendantes, n'en contient en réalité que trois, la quatrième est univoquement déterminée par les trois autres. Ceci apparaît clairement par la représentation graphique, chaque variable possède un élément de surface commun avec les trois autres. A travers ce simple exemple, nous pouvons remarquer que l'entropie de SHANNON se trouve devant l'impossibilité de donner une représentation graphique pour un tel système. En effet, elle ne peut tenir compte à la fois de l'indépendance des variables prises deux à deux, trois à trois et de leur dépendance une fois prises ensemble. Or, si l'entropie du Max rend possible la représentation graphique d'un tel système en faisant apparaître les couplages des variables prises dans leur totalité, elle possède cependant l'inconvénient de tenir compte des couplages aux niveaux inférieurs qui n'existent pas en réalité.

IV.2.6 Analyse structurale et indices de couplage

Dans le tableau initial de données, on considère chaque variable $X \in \Sigma$ comme une variable aléatoire, dont les données constituent des réalisations (au sens des probabilités). Il n'est pas nécessaire de spécifier comment les données ont été obtenues, il peut aussi bien s'agir d'un système déterministe, que d'un système stochastique. Le but de chaque analyse est l'étude des couplages entre paires ou groupes de variables. Pour mesurer les intensités de couplage entre celles-ci, il est nécessaire de faire appel à des indices (informationnels, de similarité, ...) qui permettront entre autres, la construction d'une matrice de couplage (ou graphe d'interconne-

xion). La plupart de ces indices sont issus des concepts de la théorie de l'information. En effet, celle-ci se prête bien à l'analyse des systèmes pour les raisons données au paragraphe ci-dessous.

IV.2.6.1 Pourquoi s'intéresse-t-on à la T.I. pour l'analyse des systèmes ?

Les raisons qui nous ont conduit à utiliser la T.I. sont les suivantes :

- ☛ Elle mesure des quantités d'information, sans référence aucune au contenu sémantique correspondant. Elle est donc indépendante des valeurs prises par les variables et prend en compte seulement les partitions qu'elles induisent sur l'espace d'échantillonnage Ω ;
- ☛ Elle s'applique à tout type de variables (numériques, semi-numériques, qualitatives, logiques, floues, ...) ainsi qu'à tout ensemble de modalités (structurées ou non) ;
- ☛ Elle génère des indices très sensibles à toute relation fonctionnelle (partielle ou totale) entre variables (linéaire, non linéaire, floue, logique, ...) ce qui n'est pas toujours le cas pour les indices issus par exemple de la statistique (supposant souvent des relations linéaires entre variables) ;
- ☛ Elle donne naissance à des relations dont l'interprétation intuitive est généralement naturelle et aisée [MILLER 63].

Les seules hypothèses requises en T.I. sont :

- ☛ Un nombre d'échantillons suffisamment élevé pour l'estimation des probabilités au moyen des fréquences relatives. Ceci impose naturellement un horizon d'observation assez élevé pour les systèmes dynamiques ;
- ☛ Stationnarité, ergodicité et période d'échantillonnage judicieusement choisie dans les cas des systèmes dynamiques [LARMINAT 75].

Par ailleurs, l'analyse structurale par les données qui est basée sur les concepts de la T.I. se situe à mi-chemin entre les analyses structurales physique et mathématique [DUFOUR 79]. En effet, celle-ci possède l'avantage d'utiliser les informations recueillies sur le système, reste par conséquent lié à ce système, à ces composantes (sous-systèmes), à leur structure (organisation interne). Autrement dit, elle se base sur des données statistiques et/ou informationnelles et s'écarte ainsi, de toute spéculation purement pragmatique [TORO 82].

Pour toutes ces raisons, nous allons utiliser les concepts de la théorie de l'information pour effectuer l'analyse structurale d'un système complexe. Nous allons orienter notre analyse vers les deux buts essentiels suivants :

- La modélisation : recherche des relations fonctionnelles ;
- La décomposition : recherche d'une partition du système (ou sous-systèmes).

Dans les deux cas, on a besoin d'un moyen permettant d'évaluer l'intensité de liaison entre variables ou groupes de variables. C'est pourquoi, nous allons présenter dans le paragraphe ci-dessous les différents indices de couplages généralement utilisés, nous montrerons ensuite leurs avantages et leurs limites.

IV.2.6.2 Indices de couplage

Nous envisageons successivement le cas des systèmes statiques et le cas des systèmes dynamiques.

① Cas des systèmes statiques :

Nous distinguons deux types d'indices : ceux qui sont adaptés pour une approche binaire (par paires de variables) et ceux qui s'utilisent pour une approche multidimensionnelle.

①① Analyse binaire

Les indices informationnels les plus couramment utilisés sont :

a) $d_1 = I(X : Y)$: proposée par [ASHBY 65] pour la mesure de la quantité d'information commune aux variables X et Y. Celle-ci est en réalité une mesure de couplage entre les deux variables et constitue une redondance d'information due en général à une transmission de données, ordres, paramètres, consignes, réponses, ... à travers un système physique (canal ou autre) allant de X vers Y et/ou de Y vers X.

b) $d_2 = I(X : Y) / H(Y)$: [CONANT 72] : est un coefficient de couplage normalisé. En effet : $0 \leq d_2 \leq 1$; $d_2 = 1$ ssi il existe une application de M_X dans M_Y où M_X et M_Y sont respectivement les ensembles de modalités de X et de Y. L'auteur propose cet indice pour la mise en évidence des structures du système et particulièrement les sous-systèmes faiblement couplés.

c)

$$d_3 = \frac{I(X : Y)}{\sqrt{H(X) \cdot H(Y)}}$$

[RICHETIN 75] : proposé dans le but d'éliminer l'influence du nombre de valeurs (ou modalités) possibles de chacune des variables. Comme l'indice précédent, celui-ci est également normalisé puisque : $0 \leq d_3 \leq 1$: $d_3 = 1$ ssi il existe une application biunivoque de M_X dans M_Y .

d)

$$d_4 = \frac{H(X/Y) + H(Y/X)}{H(X \cup Y)}$$

[LOPEZ DE MANTARAS 77] : Cet indice n'est autre que le cardinal de la différence symétrique normalisé.

e)

$$d_5 = \frac{I(X:Y)}{H(X \cup Y)}$$

[DUSSAUCHOY 80] : Il n'est autre que le complément à 1 de l'indice précédent. L'auteur démontre que cet indice de couplage est une relation de similarité floue transitive.

f)

$$d_6 = \frac{I(X:Y)}{\min [H(X), H(Y)]}$$

[YASUICHI HORBIBE 85] : Celui-ci montre que cet indice est une distance et peut être utilisé pour la mesure de la corrélation entre les variables du système.

Il existe dans la littérature consacrée à l'analyse structurale, d'autres indices qui sont pour la plupart très proches de ceux donnés ci-dessus.

En définitive, les indices proposés ont pour but d'effectuer une analyse structurale (décomposition, modélisation, identification, ...). Or, si on se réfère aux méthodes de décomposition hiérarchique, de classification automatique, on s'aperçoit que de telles approches utilisent des indices calculés à partir des tableaux de similarité (ou de dissimilarité) extraits du tableau initial de données. De tels indices sont appelés généralement indices de similarité (ou de dissimilarité) [LERMAN 76]. Une analogie (ou un parallèle) peut être alors faite entre les indices informationnels et les indices de similarité. En effet, une variable dichotomique induit une partition de Ω en deux classes. La similarité entre deux variables dichotomiques est définie de façon classique à partir de l'intersection de deux partitions, en utilisant comme mesure le cardinal des classes de cette intersection qui ne constitue en réalité qu'un exemple de semi-évaluation [ABID 79]. Par analogie avec les indices de similarité classique et à l'aide de l'interprétation graphique de l'entropie, il est alors possible de définir l'ensemble des indices de similarité "informationnels" correspondants. En effet, dans le premier cas, on définit les quantités s , t , u , v , t et T [fig. I.6] (pour plus de détail voir LERMAN 76), alors que dans le deuxième cas, on fait apparaître les quantités : $I(X:Y)$, $H(X/Y)$, $H(Y/X)$, $H(\Sigma/X \cup Y)$ et $H(\Sigma)$ [fig. I.7]

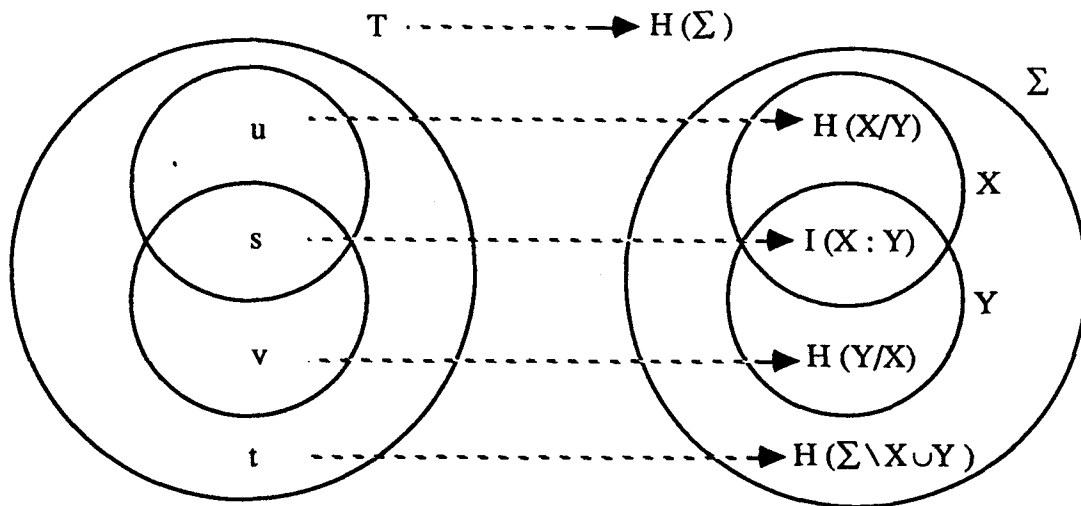


Fig. I.6

Fig. I.7

L'analogie entre les deux représentations nous conduit alors, à partir des indices de similarité (ou de dissimilarité) classiques, de retrouver la plupart des indices "informationnels" les plus connus et proposer certains qui sont nouveaux.

AUTEURS	INDICE DE SIMILARITÉ	INDICE "INFORMATIONNEL"	
JACCARD	$\frac{s}{s + u + v}$	$\frac{I(X : Y)}{H(X \cup Y)}$	DUSSAUCHOY 80
DICE	$\frac{2s}{2s + u + v}$	$\frac{2I(X : Y)}{H(X) + H(Y)}$	indice de distance
KENDALL & COL	$1 - \frac{u + v}{T}$	$1 - \frac{H(Y/X) + H(X/Y)}{H(\Sigma)}$	le complément à 1 de cet indice est une distance
SOKAL ET SNEATH	$\frac{s}{s + 2(u + v)}$	$\frac{I(X : Y)}{H(\Sigma) + H(X/Y) + H(Y/X)}$	indice de distance

KULEZYNSKI	$\frac{s}{u+v}$	$\frac{I(X:Y)}{H(Y/X) + H(X/Y)}$	indice de distance
KULEZYNSKI	$\frac{1}{2} \left[\frac{s}{s+u} + \frac{s}{s+v} \right]$	$\frac{1}{2} \left[\frac{I(X:Y)}{H(X)} + \frac{I(X:Y)}{H(Y)} \right]$	forme symétrique de l'indice de CONANT (1972)
OCHIAI	$\frac{s}{\sqrt{(s+u)(s+v)}}$	$\frac{I(X:Y)}{\sqrt{H(X)H(Y)}}$	[RICHETIN 75]
RUSSEL ET RAO	$\frac{s}{T}$	$\frac{I(X:Y)}{H(\Sigma)}$	le complément à 1 de cet indice est une distance
CARDINAL DE LA DIFFERENCE SYMETRIQUE	$u+v$	$H(Y/X) + H(X/Y)$	LOPEZ 77
	$\frac{u+v}{u+v+s}$	$\frac{H(Y/X) + H(X/Y)}{H(X \cup Y)}$	LOPEZ 77
	$\frac{u+v}{u+v+t}$	$\frac{H(Y/X) + H(X/Y)}{H(\Sigma) - I(X:Y)}$	c'est une distance

Remarque :

L'ensemble de ces indices servent à définir la ressemblance entre éléments pris deux à deux définis sur Ω (indices de similarité) ou sur Σ (indices informationnels). En effet, la notion de similarité rejoint celle de distance dans son utilisation, elles constituent ensemble un instrument de mesure d'écart bien qu'elles aient des définitions antinomiques [BOURTON 75]. En réalité, l'application et le choix entre les deux formalismes dépendent du type de système et de la nature des données à traiter (données logiques, données quantitatives, ...). Par ailleurs, certaines approches classiques font appel à :

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} \quad \text{coefficient de corrélation linéaire des variables X et Y}$$

$$\text{avec } \sigma(X) = \sqrt{\text{var}(X)} \quad \text{et} \quad \sigma(Y) = \sqrt{\text{var}(Y)}$$

S'il existe une relation linéaire entre X et Y de la forme $Y = aX + b$, on montre alors que : $\text{var}(Y) = |a| \sigma(X)$ et $\text{cov}(X, Y) = a \text{var}(X)$ d'où :

$$\rho(X, Y) = \begin{cases} 1 & \text{si } a > 0 \\ -1 & \text{si } a < 0 \end{cases}$$

Si X et Y sont deux variables gaussiennes centrées telles que :

$E(X) = 0, E(Y) = 0, E(X^2) = \sigma^2(X), E(Y^2) = \sigma^2(Y)$ et $E(X.Y) = \text{cov}(X, Y)$ (E : espérance mathématique, on montre alors [DUFOUR 79] que :

$$I(X : Y) = -\text{Ln} \sqrt{1 - [\rho(X, Y)]^2} \quad (\text{I. 70})$$

résultat classique qui donne la relation entre le coefficient de transinformation et le coefficient de corrélation linéaire dans le cas gaussien. Il faut cependant signaler que la nullité du coefficient de corrélation signifie uniquement l'absence de relations linéaires mais n'exclut pas l'existence d'autres relations. Ainsi, par exemple, X et X² ou sin X et cos X sont deux fonctions qui sont sans contexte dépendantes mais présentent un coefficient de corrélation linéaire nul.

①② Analyse multidimensionnelle

Si l'analyse purement binaire d'un système est facile à mettre en œuvre au moyen d'indices de similarité, de relations binaires, de graphes valués ou non, ... elle possède par contre des inconvénients qui résident dans le fait que celle-ci peut laisser de côté certains couplages du système qui ne peuvent être appréhendés que par une approche multidimensionnelle. Dans ce cas, on fait appel aux indices informationnels suivants :

a) $D_1 = I(X_1 : X_2 : \dots : X_N)$: la transinformation totale dans un système proposée par ASHBY [1965] qui est considérée comme le paramètre le plus important de celui-ci. En effet, il mesure la quantité totale de relation que l'on peut espérer extraire du système décrit par les variables X_1, X_2, \dots, X_N . C'est seulement lorsqu'il existe des relations (présence des redondances) qu'on pourra en faire une description simplifiée. Par ailleurs, dans le cas où une structure complexe ne contient pas de redondance, elle sera elle-même sa plus simple description.

b)

$$D_2 = \frac{I(S_1 : S_2 : \dots : S_k)}{I(X_1 : X_2 : \dots : X_N)}$$

[CONANT 72] [RICHETIN 75] mesure la qualité d'une partition du système en k classes. Si l'objectif fixé est la décomposition du système en sous-systèmes faiblement couplés, une telle décomposition sera d'autant plus satisfaisante que D_2 est faible. CONANT fut le premier à proposer l'utilisation d'un tel indice comme critère d'optimalité pour juger une partition d'un système. La partition optimale doit nécessairement conduire à une valeur de D_2 minimale. Dans ce sens, TORO [1982] propose la minimisation de D_2 à tous les niveaux du treillis $(P(\Sigma), \leq)$ ce qui conduit à l'obtention des optimums locaux vérifiant la condition d'optimalité.

c) $D_3 = I_{\Sigma_{X \cup Y}}(X : Y)$: est une mesure de couplage direct entre X et Y . Le calcul de cet indice nécessite un plus grand nombre d'observations que celui de $I(X : Y)$. Il fait intervenir le calcul de l'information entre X et Y pour chaque réalisation de $\Sigma\{X \cup Y\}$ [RICHETIN 75].

② Cas des systèmes dynamiques :

L'étude des couplages d'un système dynamique, nécessite la prise en compte des valeurs de chacune de ces variables à instant quelconque et certaines de leurs valeurs à des instants antérieurs. Si on note Y' la variable Y à un instant décalé d'une période d'échantillonnage T par rapport à l'instant de la mesure de X , les principaux indices de couplage dynamique peuvent être résumer comme suit :

a) $\delta_1 = I(X : Y')$ mesure le couplage dynamique entre X et Y dans le sens $X \rightarrow Y$. Dans le cas général, cet indice n'a pas la propriété de symétrie. Le couplage $Y \rightarrow X$ mesuré par $I(Y : X')$ est différent de $I(X : Y')$.

b)

$$\delta_2 = \frac{I(X : Y')}{\sqrt{H(X) \cdot H(Y')}}$$

[RICHETIN 75] : indice normalisé puisque : $0 \leq \delta_2 \leq 1$. Si $\delta_2 = 1 \Leftrightarrow$ il existe une application biunivoque de M_X dans $M_{Y'}$. Cette condition nécessaire et suffisante est très restrictive quant à la forme de liaison entre X et Y' [RICHETIN 75]. C'est pourquoi celui-ci préfère l'utilisation de l'indice suivant :

c)

$$\delta_3 = \frac{I(X : Y')}{H(Y')}$$

[CONANT 69] : cet indice est plus général que le précédent puisque $\delta_3 = 1$ lorsqu'il existe simplement une application de M_X dans $M_{Y'}$. Comme dans le cas statique, un tel coeffi-

cient est normalisé dans le but d'éliminer en partie l'influence du nombre de classes [DUFOUR 79].

d)

$$\delta_4 = \frac{I_{\Sigma X}(X : Y')}{H(Y' / \Sigma X)}$$

[RINK 73] : cet indice mesure le couplage dynamique direct entre X et Y. En effet, comme le signale l'auteur, ce coefficient permet l'élimination des couplages dynamiques apparents provenant des couplages statiques. Puisque, même si $I_{\Sigma X}(X : Y') = 0$ on peut avoir $I(X : Y') \neq 0$ si $I(Z : X) \neq 0$ et $I(X : Y) \neq 0$

Un tel coefficient n'est pas facile à mettre en œuvre puisqu'il nécessite beaucoup d'échantillons pour que le résultat soit significatif. Une telle procédure apparaît beaucoup plus longue que celle qui utilise $I(X : Y')$.

e)

$$\delta_5 = \frac{I(X : Y')}{H(X \cup Y')}$$

[CONANT 72, DUFOUR 78] : est un indice de couplage dynamique normalisé, non symétrique. Le complément à 1 de cet indice constitue un indice de dissimilarité dynamique [PRADIN 86]. Celui-ci constate qu'un tel indice met bien en évidence les couplages dynamiques réels et qu'il est également sensible aux couplages dynamiques apparents induits par des couplages statiques à travers d'autres variables du système.

Toutes les interprétations formulées dans le cas statique restent valables dans le cas dynamique à condition que la période d'échantillonnage soit compatible avec la dynamique du système. Or, en pratique une telle hypothèse est difficilement vérifiable, ceci est d'autant plus vrai que le système est composé de plusieurs sous-systèmes de dynamiques différentes.

IV.2.7 Problèmes posés par les indices de couplage en présence du bruit : approche globale et approche locale

IV.2.7.1 Position du problème

Toute analyse par les données repose sur deux opérations successives : dans un premier temps, le recueil des données et ensuite leur analyse pour atteindre l'objectif visé. Il est clair que la première opération est sans conteste la plus importante. A la limite, on peut dire que l'analyse est purement mécanique lorsqu'on dispose des méthodes appropriées alors que le recueil suppose jugement, choix, discernement et bon sens. En effet, la transmission d'un messa-

ge à travers un système (canal ou autre), la mesure d'une grandeur à l'aide d'un capteur, la réponse à un questionnaire, ... sont sujettes à des perturbations d'origines diverses : bruit, aléas, erreurs ... Par conséquent, la quantité d'information véhiculée est soumise à un facteur de dégradation qui se traduit par une perte d'information. L'effet d'une telle dégradation peut être important ou négligeable. En effet, comme le dit OSWALD [1986] à propos de la transmission d'un message "dans de nombreux cas, la restitution incorrecte d'une lettre dans un message télégraphique n'est pas très gênante parce que le langage est très redondant et que le destinataire corrige sans effort le texte qu'il reçoit. S'il s'agit d'une séquence de données numériques, destinée à un ordinateur, une erreur sur un mot de la séquence peut avoir au contraire des effets catastrophiques". De même, dans un système technologique, un capteur défectueux donne une grandeur erronée qui a pour conséquence un déplacement du point de fonctionnement qui peut se traduire par un changement du mode de fonctionnement d'un tel système. Ainsi, par exemple, dans une station d'épuration, il existe trois modes de fonctionnement relatifs aux faibles charges, charges moyennes et fortes charges (ou charges toxiques). Si la concentration d'entrée en substrat est sur-estimée, alors ceci peut se traduire par un fonctionnement fortes-charges de la station qui a pour conséquence une perte d'énergie inutile au niveau de l'oxygénation du bassin d'aération. De même, une sous-estimation d'une telle concentration a pour effet soit une augmentation du temps de séjour de l'eau dans le bassin ou bien une concentration résiduelle importante en sortie.

En définitive, une bonne analyse (ou un bon fonctionnement d'un système) ne peut être obtenue que si les deux conditions suivantes sont vérifiées simultanément.

i) Les mesures effectuées sur la population d'apprentissage (ou système) doivent être représentatives, traitées et filtrées. Il est bien évident qu'une analyse effectuée sur des données "fausses" ou "trop bruitées" n'aura aucune valeur et ne peut aboutir aux résultats escomptés.

ii) Disposer d'indices permettant de juger réellement les quantités d'informations véhiculées par le système. Ainsi, le choix d'un mauvais indice peut masquer, sur-évaluer ou minimiser le flux informationnel circulant à l'intérieur et/ou entre les sous-systèmes.

Si les deux hypothèses sont vérifiées, il est possible de procéder à l'analyse du système et on peut alors espérer, en utilisant au mieux ses caractéristiques structurales, mettre à jour ses redondances ou son organisation interne.

IV.2.7.2 Formulation du problème

On désire analyser un système et on suppose qu'il a été fixé d'avance une division de ses variables en deux groupes : les variables à expliquer et les variables explicatives.

- Soient X et Y deux ensembles de ces variables, entre lesquels on désire analyser les relations. Celles-ci peuvent être des variables vectorielles ou non.

- $M_X = \{\alpha_i, i = 1, \dots, n\}$ et $M_Y = \{\beta_j, j = 1, \dots, m\}$: les ensembles de modalité de X et de Y.

- Les variables X et Y génèrent un tableau de contingence $[P_{ij}]$ dans lequel la case (i,j) contient P_{ij} la probabilité d'occurrence conjointe des modalités α_i et β_j . Analysons les couplages entre X et Y sous les deux hypothèses suivantes :

* Hypothèse H_0 : liaison fonctionnelle bruitée entre X et Y

$$\text{Soit } P_{j/i} = \max_{k \neq j} P_{k/i}, \quad k = 1, \dots, m$$

Supposons que $P_{j/i} \gg P_{k/i}, \forall k \neq j$, autrement dit la ligne i est constituée d'un élément correspondant à un phénomène réel auquel se superpose un phénomène purement aléatoire que l'on peut modéliser par une densité uniforme. Dans ce cas, on peut alors écrire :

$$H(Y/X = \alpha_i) = -P_{j/i} \text{Log } P_{j/i} - (1 - P_{j/i}) \text{Log } \frac{1 - P_{j/i}}{m - 1}$$

Supposons par ailleurs, que les n lignes du tableau de contingence ont les mêmes probabilités marginales et présentent les mêmes distributions où le phénomène réel est sur la diagonale et que le bruit se répartit uniformément de part et d'autre de celle-ci. D'où $H(X) = \text{Log } n$, $H(Y) = \text{Log } m$ et $H(Y/X) = H(Y/X = \alpha_i)$.

Le calcul de $I(X : Y)$ sous cette hypothèse donne :

$$I(X : Y)_0 = H(Y)_0 - H(Y/X)_0 = \text{Log } m + P_{j/i} \text{Log } P_{j/i} + (1 - P_{j/i}) \text{Log } \frac{1 - P_{j/i}}{m - 1} \quad (\text{I.71})$$

* Hypothèse H_1 : Liaison fonctionnelle partielle non bruitée entre X et Y.

Supposons maintenant que chaque ligne contient seulement deux éléments représentant chacun un phénomène réel de probabilités respectives : $P_{j/i}$ et $1 - P_{j/i}$ avec $P_{j/i}$ identique au cas précédent et

$$(1 - P_{j/i}) = \sum_{k \neq j} P_{k/i} \quad (\text{sous } H_0)$$

L'entropie d'une distribution ligne s'écrit dans ce cas :

$$H(Y/X = \alpha_i) = - P_{ji} \text{Log } P_{ji} - (1 - P_{ji}) \text{Log } (1 - P_{ji})$$

En se plaçant dans les mêmes conditions que précédemment, c'est-à-dire : $H(X) = \text{Log } n$, $H(Y) = \text{Log } m$ et $H(Y/X = \alpha_i) = H(Y/X)$, le calcul de $I(X : Y)$ sous cette hypothèse donne :

$$I(X : Y)_1 = H(Y)_1 - H(Y/X)_1 = \text{Log } m + P_{ji} \text{Log } P_{ji} + (1 - P_{ji}) \text{Log } (1 - P_{ji}) \quad (\text{I.72})$$

* Conséquence :

$$I(X : Y)_0 = I(X : Y)_1 - (1 - P_{ji}) \text{Log } (m - 1) \quad (\text{I.73})$$

d'où $I(X : Y)_1 > I(X : Y)_0$: ce qui signifie que l'information échangée entre X et Y est plus importante sous l'hypothèse H_1 que sous l'hypothèse H_0 . Or en réalité, il existe une relation fonctionnelle entre X et Y, sous H_0 , qui n'a pas été décelée à tort à cause de la présence du bruit alors que manifestement, il n'existe qu'une liaison partielle entre X et Y sous H_1 . Par conséquent, une analyse basée sur l'utilisation d'un indice tel que $I(X : Y)$ sur des données bruitées peut conduire à :

- une décomposition non optimale du système ;
- à des erreurs de modélisation ... etc.

IV.2.7.3 Approche globale, approche locale

a) Approche globale

L'approche globale consiste à effectuer l'analyse sur l'ensemble d'échantillonnage Ω , sur l'ensemble des parties et l'ensemble des partitions induites sur Ω . Quelque soit l'analyse menée (méthodes géométriques de l'A.D., méthodes informationnelles, ...), elle ne peut procéder que globalement par la mise en œuvre d'outils tels que les indices précédemment décrits. Une telle analyse, ne peut mesurer que la quantité moyenne de relations, de lois, du flux informationnel entre les différentes composantes du système. Or, la présence du bruit sur certaines modalités du système, peut dégrader l'information véhiculée par le système qui se traduit par une mauvaise appréhension des relations de dépendance ou de causalité entre les variables et groupes de variables de celui-ci. De telles modalités ne sont pas détectées par une approche globale, si bien que par exemple, deux variables fortement liées peuvent se trouver dans des sous-systèmes différents. Autrement dit, l'utilisation des indices précédents directement sur les données brutes, au moyen d'une approche globale, s'avère insuffisante et peut même induire l'analyste en erreur. Ceux-ci ne tiennent pas compte en effet, de la nature de l'information recueillie au moyen des capteurs, ..., et ne peuvent par conséquent distinguer entre une information issue d'un phénomène déterministe et celle provenant d'un comportement purement aléatoire. D'où l'inconvénient d'une approche globale du système. On peut remarquer toutefois que

celle-ci reste valable sous l'hypothèse que le phénomène aléatoire se répartit de la même façon sur l'ensemble des modalités du système, de telle sorte que les variables dépendantes, même bruitées peuvent se trouver dans le même sous-système. D'un point de vue opérationnel, une telle hypothèse est non réaliste étant donné que les protocoles de mesure sont de natures différentes, de précision et de qualité non équivalentes, certains sont fiables, d'autres le sont moins ...

b) Approche locale

Celle-ci opère sur les parties de Ω , autrement dit sur les classes de l'ensemble des partitions $P(\Omega)$ prises séparément. En effet, l'approche locale consiste à travailler sur un sous-échantillon de la population Ω , de dimension plus faible, donnant naissance à sous tableau de données. L'analyse de ce dernier, en tenant compte de la nature de l'information qu'il recèle nous conduira à séparer celle-ci en :

- une information correspondant à un phénomène réel et
- une information provenant d'un phénomène purement aléatoire.

Nous montrerons aux chapitres II et III comment il est possible d'extraire l'information réellement représentative du sous-échantillon analysé. Pour l'instant, signalons qu'une analyse locale effectuée sur un ensemble de sous-échantillons permet de concentrer les moyens de modélisation, d'identification, ... sur des sous-systèmes de taille réduite. En effet, celle-ci mesure la quantité de lois à l'intérieur des sous-systèmes "lois régionales". Nous montrerons également que l'application des approches locales conduit à faciliter la modélisation du fonctionnement d'une station d'épuration des eaux usées, en distinguant trois modèles du système correspondant aux trois modes de fonctionnement de celui-ci. Par ailleurs, si dans un système technologique, un capteur défaillant ou en mauvais état de fonctionnement fournit une information dégradée, dans une approche globale, celle-ci peut être masquée alors que dans une approche locale elle peut être détectée et éventuellement corrigée.

V - CONCLUSION

Nous avons montré, dans ce chapitre, que l'analyse des systèmes complexes sur lesquels on ne possède aucune information a priori, peut être abordée sous deux angles différents : d'une part, les méthodes géométriques de l'analyse des données, d'autre part, les concepts issus de la théorie de l'information. Nous avons plus particulièrement présenté les principaux résultats de cette dernière approche que nous allons utiliser dans la suite de ce travail. En effet, nous avons fait appel aux différents concepts algébriques de la théorie de l'information susceptibles d'être appliqués en analyse structurale. Par souci de clarté, tous les indices et relations

issus de ces concepts sont interprétés intuitivement en terme d'information. L'utilisation de l'entropie de Max pour certains types de systèmes, permet d'appréhender les couplages binaires qui ne peuvent être décelés par une autre entropie. Elle conduit également, dans certains cas, à la représentation graphique du système. Par ailleurs, les indices de similarité classiques et les indices informationnels conçus indépendamment les uns des autres, malgré la similitude des démarches, peuvent être harmonisés sous un même formalisme. De nombreux indices sont alors présentés. La présence du bruit sur le système, limite en partie une utilisation directement sur les données brutes recueillies sur celui-ci. C'est pourquoi, nous allons montrer au chapitre suivant, comment de tels indices peuvent être appliqués pour effectuer une analyse structurale du système qui sera orientée essentiellement vers les buts suivants :

- modélisation : recherche des relations fonctionnelles ;
- décomposition : recherche des systèmes faiblement couplés.

CHAPITRE II

MODELISATION STRUCTURALE DES
SYSTEMES COMPLEXES
RECHERCHE DE RELATIONS
FONCTIONNELLES

I - INTRODUCTION

L'analyse structurale répond à la nécessité de disposer d'outils et de méthodes, qui face à une masse importante de données, permettent d'en extraire les variables pertinentes et leurs relations en vue d'une modélisation du système. Fondée sur le concept d'"explication" d'un ensemble de variables au moyen d'un autre, cette approche doit fournir des techniques suffisamment générales, compte tenu du fait que les domaines d'application de "l'approche systèmes" ne cessent de s'élargir. Ainsi les approches statistiques d'analyse des données, d'identification se trouvent parfois limitées par les hypothèses restrictives, quelquefois difficilement vérifiables qu'elles supposent (variables numériques, relations linéaires, ...). La modélisation structurale utilisant les concepts de la théorie de l'information présente l'avantage de s'appliquer à tout type de variables (quantitatives, qualitatives, à ensemble de modalités structuré ou non) ainsi qu'à tout type de relation (linéaire, non linéaire, floue, codage, ...). Le premier qui a proposé des travaux dans ce sens fut à notre connaissance P.C. CONANT [1972]. Ces travaux ont été poursuivis par R.E. RINK [1973], M. RICHTIN [1975], Z. ABID [1979], J. DUFOUR [1979] et plus récemment V. TORO [1981,1982] et M. STAROSWIECKI [1981,1984] qui ont proposé quelques algorithmes "d'explication". Cependant, ces travaux se trouvent limités lorsqu'il s'agit de les mettre en œuvre sur des systèmes réels soumis à des perturbations de natures diverses.

Dans ce chapitre, nous présenterons, quelques concepts de base exposés par ces auteurs et nous montrerons leurs limites. Nous développerons ensuite un algorithme de recherche des sous-ensembles explicatifs. Son caractère exhaustif garantit l'optimalité des solutions trouvées. On met également en évidence les notions de degrés de liberté d'un système, de sous-ensembles explicatifs multiples d'une variable, de relations redondantes. Par ailleurs, pour des systèmes suffisamment complexes, l'expérience prouve qu'il est en général difficile d'expliquer une variable de façon satisfaisante sur l'ensemble de ses modalités surtout en présence du bruit. Dans ce sens, nous considérons une partition de cet ensemble en un sous-ensemble de modalités bien expliquées et un sous-ensemble de modalités dont l'explication requiert l'introduction de variables explicatives supplémentaires. Nous présentons dans ce cas un algorithme de recherche de ces sous-ensembles explicatifs supplémentaires.

Enfin nous développons une application de la modélisation structurale au domaine de l'épuration des eaux usées utilisant la voie biologique.

II - ANALYSE INFORMATIONNELLE ET MODELISATION

II.1 Position du problème

La complexité des systèmes pris en compte actuellement augmente de plus en plus et pose de sérieux problèmes de modélisation, dûs au grand nombre de variables nécessaires à une description complète, ainsi qu'à leurs natures différentes (quantitatives, qualitatives, floues, ...). Une voie possible pour diminuer la complexité de ces problèmes consiste à utiliser au mieux les caractéristiques structurales des systèmes étudiés. La détermination de celles-ci, conduisant à la mise en évidence de certains sous-systèmes du système global et de leur structure d'interconnexion est donc une étape importante de l'analyse [STAROSWIECKI 1983]. La modélisation structurale constitue dans ce sens une approche permettant à la fois :

- de décomposer le système en sous-systèmes indépendants ou non ;
- de mettre à jour les redondances, c'est-à-dire l'organisation interne du système [WALLISER 1977].

Lorsque le graphe relationnel est connu, il suffit de déterminer sa structure hiérarchisée pour obtenir celle du système [RICHETIN 1975, MILGRAM 1975]. Dans le cas contraire, l'observation sur un horizon donné des variables du système conduit à l'étude de leurs couplages. Cette démarche repose sur une tentative d'explication de chaque variable du système au moyen des autres variables (recherche des redondances).

Soit, $\Sigma = \{X_i, i = 1, \dots, N\}$ l'ensemble des variables nécessaires à la description du système. Dans le cas d'un système dynamique, on suppose que l'explication des variables à l'instant présent, nécessite la connaissance de l'information fournie par les variables aux instants précédents. Chaque variable $X_i \in \Sigma^D$ (Σ^D système dynamique) est un vecteur à p composantes et s'écrit :

$$X_i = (x_i(t), x_i(t-1), x_i(t-2), \dots, x_i(t-p+1))$$

Autrement dit X_i regroupe les p dernières réalisations de la variable x_i . Ainsi Σ^D contient toutes les variables $X_i, i = 1, \dots, N$, à l'instant présent décalées d'une période, décalées de deux périodes, ..., décalées de $(p-1)$ périodes.

L'objectif visé par cette analyse est de chercher dans quelle mesure on peut établir un modèle du système sous forme de relations fonctionnelles.

A cet effet, nous allons montrer dans un premier temps l'insuffisance des approches classiques.

II.2 Insuffisance des approches classiques

On cherche à établir un modèle du système. Ceci revient à déterminer la structure de liaison entre variables ou groupes de variables du système. En effet, il s'agit de chercher dans quelle mesure chaque variable du système peut être expliquée (ou modélisée) au moyen d'un groupe d'autres variables du système. Une approche classique d'explication s'effectue généralement de deux manières différentes : la première fait apparaître les couplages de l'ensemble des variables pris deux à deux (analyse binaire), par contre la deuxième approche met en évidence les couplages entre variables et groupes de variables (ou sous-systèmes) (analyse multi-dimensionnelle). Dans le premier cas, des indices sont définis (cf. chapitre I, indices d_1 à d_6 et δ_1 à δ_4), des relations binaires entre variables sont alors déduites soit sous forme de graphe valué ou non [MALOUM 1980], ou sous forme de relations de ressemblance, de similarités ou de distances [PRADIN 1984]

Si une telle approche est facile à mettre en œuvre, à cause de sa simplicité, ses conclusions peuvent être par contre parfois contestées puisqu'elles laissent échapper certains couplages importants du système. Dans le deuxième cas, on fait appel à un indice qui tient compte de l'aspect multi-dimensionnel de l'analyse (cf. chapitre I, indice D_3 et S_4) mais ne conduit pas forcément à une solution du problème, d'autant plus que sa mise en œuvre nécessite malheureusement un temps de calcul prohibitif qui a pour conséquence une diminution de la capacité d'une telle approche à traiter des systèmes de grande dimension.

II.2.1 Insuffisance de l'analyse binaire

Si on se réfère à l'exemple traité au chapitre I (tableau de données fig. I.4, paragraphe IV.2.5), ou plus généralement si on considère un système logique combinatoire composé de N variables où chacune d'elles est fonction des parités des $(N - 1)$ autres, on montre alors, dans le cas de l'utilisation de l'entropie de SHANNON, que : l'information échangée entre les N variables, prises ensemble, est positive, alors qu'elle est nulle pour tout sous-ensemble de dimension inférieure à N et particulièrement entre paires de variables. Autrement dit :

$$I(X_1 : X_2 : \dots : X_N) > 0 \quad \text{mais} \quad \forall X_i, X_j \in \Sigma : I(X_i : X_j) = 0, \quad \forall i \neq j \quad (\text{II.1})$$

Une analyse basée sur l'utilisation des indices binaires peut conclure hâtivement qu'il n'y a aucun lien entre les variables du système alors qu'en réalité, il existe un lien qui n'a pas été pris en compte par une telle analyse. Ceci apparaît clairement lors du calcul qui donne

$$H(X_i / \Sigma \setminus X_i) = 0, \quad \forall i \in \{1, \dots, N\},$$

résultat qui exprime le fait que chaque variable est totalement déterminée lorsque les $(N - 1)$ autres sont connues.

Pour résoudre ce problème, on peut effectuer une analyse multi-dimensionnelle. Or,

cette dernière approche, comme nous l'avons signalé plus haut, peut s'avérer insuffisante.

II.2.2 Insuffisance de l'analyse multidimensionnelle

Pour mesurer le degré de liaison entre deux variables du système, on fait appel dans ce cas à la transformation directe [indices D_3 et S_4 chapitre I, paragraphe IV.2.6.2.] ou conditionnelle dans le but de déceler les couplages ignorés ou laissés de côté par une analyse purement binaire. Or celle-ci possède deux inconvénients majeurs :

i) soient 2 ensembles explicatifs disjoints S_1 et $S_2 \in \mathcal{P}(\Sigma)$ d'une variable Y réalisant une partition de Σ en 3 classes telles que :

$$\Sigma = Y \cup S_1 \cup S_2 \quad \text{et} \quad Y \cap S_1 \cap S_2 = \emptyset$$

Théorème 222 : si $I(Y : S_1) = I(Y : S_2) = I(Y : S_1 \cup S_2)$ alors

$$I_{\Sigma \setminus Y \cup S_1}(Y : S_1) = I_{\Sigma \setminus Y \cup S_2}(Y : S_2) = 0$$

Preuve : $I_{S_2}(Y : S_1) = H(Y/S_2) - H(Y/S_1 \cup S_2)$
 $= [H(Y \cup S_2) - H(S_2) - H(Y \cup S_1 \cup S_2) + H(S_1 \cup S_2)] + H(Y) - H(Y)$
 $= I(Y : S_1 \cup S_2) - I(Y : S_2) = 0 \quad \text{F.D.}$

Interprétation : L'hypothèse signifie qu'il existe une redondance entre les trois composantes du système. En effet,

$$I(Y : S_1) = I(Y : S_2) = I(Y : S_1 \cup S_2) \Rightarrow H(Y/S_1) = H(Y/S_2) = H(Y/S_1 \cup S_2),$$

donc il existe une redondance, puisque S_1 et S_2 fournissent la même quantité d'information au sujet de Y . Celle-ci ne peut être mise en évidence par la transformation directe.

ii) La détermination d'un tel indice, nécessite à chaque fois le calcul de 4 entropies conjointes, un calcul qui risque d'être long et fastidieux dans le cas d'un système de dimension importante.

II.3 Concept de modélisabilité

Le but visé par cette procédure est de trouver dans quelle mesure une variable Y est liée (dépendante, fonction, attachée, ...) à un ou plusieurs sous-ensembles de variables S ,

$$S \in \mathcal{P}(\Sigma \setminus Y).$$

Si l'on désire expliquer une telle variable au moyen d'un sous-ensemble d'autres variables, il convient de choisir un sous-ensemble apportant l'information maximale sur la variable à

expliquer, ou encore telle que l'entropie résiduelle de celle-ci sachant le sous-ensemble explicatif soit la plus faible possible.

Soit $Y \in \Sigma$ et $S \in \mathcal{P}(\Sigma \setminus Y)$ on définit :

. $H(Y/S)$: l'entropie de Y non expliquée par S

ou encore sous forme normalisée :

. $t(Y,S) = H(Y/S) / H(Y)$: fraction de l'entropie de Y non expliquée de S .

Ces indices permettent de définir la modélisabilité d'une variable à partir des autres variables retenues pour décrire le système. Ainsi, le plus grand ensemble susceptible d'expliquer Y étant $\Sigma \setminus Y = \theta$ et $H(Y/\Sigma \setminus Y)$ la partie de l'entropie de Y non explicable par $\Sigma \setminus Y$, une faible valeur de $t(Y,\theta)$ indique qu'il est possible d'établir une relation du type $Y \cong f(\theta)$, alors que dans le cas contraire où $t(Y,\theta)$ est non négligeable, cela signifie que certaines variables importantes n'ont pas été prises en compte ou qu'il n'est pas possible de construire un modèle déterministe.

La modélisation structurale permet alors de mettre en évidence les relations existant entre les variables du système, à partir de l'idée qu'une relation se traduit par une transmission d'information. Ainsi, minimiser $H(Y/S)$ revient à maximiser $I(Y : S)$ puisque :

$$I(Y : S) + H(Y/S) = H(Y) = \text{constante}$$

Pour éviter les inconvénients d'une analyse purement binaire, nous proposons une démarche descendante pour la recherche des sous-ensembles explicatifs de chaque variable. Une telle démarche commence par opérer au niveau N , puis au niveau $(N - 1)$, et de proche en proche jusqu'à ce que la variable à expliquer présente une entropie résiduelle minimale.

III - MODELISATION DANS LE CAS D'UN SYSTEME DETERMINISTE : DETERMINATION DES SOUS ENSEMBLES EXPLICATIFS

III.1 Formulation du problème

A chaque X_i , on associe l'ensemble $\theta_i = \Sigma \setminus X_i$. Il s'agit de trouver pour chaque variable $X_i \in \Sigma$ le sous-ensemble (ou les sous-ensembles) explicatif $S_i \in \mathcal{P}(\Sigma \setminus X_i)$ optimal. Celui-ci est tel que :

$$H(X_i/S_i) \leq H(X_i/S), \quad \forall S \in \mathcal{P}(\Sigma \setminus X_i) \quad (\text{II.3})$$

Cette façon de procéder conduit à la solution triviale $S_i = \Sigma \setminus X_i$ qui ne présente pas beaucoup d'intérêt puisqu'elle conclut que chaque variable doit être expliquée par les $(N - 1)$ autres variables du système. Or notre but, est de trouver par chaque variable m (ou ses) ensemble explicatif optimal, de cardinal le plus faible possible pour faciliter les étapes ultérieures de l'analyse : modélisation, identification, commande... Compte tenu de la monotonie de $H(X_i / S_j)$ par rapport à l'inclusion dans $\mathcal{P}(\Sigma \setminus X_i)$, il convient de chercher les optimums locaux par niveaux. A chaque niveau k ($1 \leq k \leq |\theta_i|$), l'ensemble explicatif noté S_{ij}^k peut être unique ou non étant donné que dans un système de grande dimension une variable peut être liée à une ou plusieurs groupes de variables par des relations fonctionnelles différentes. Soit S_{ij}^k un ensemble explicatif de X_i au niveau k . Le processus de recherche de ces sous-ensembles explicatifs peut être énoncé comme suit :

A chaque $X_i \in \Sigma$, $i = 1, \dots, N$ et pour chaque $k = |\theta_i|, |\theta_i| - 1, \dots, 2, 1$, trouver les sous-ensembles explicatifs S_{ij}^k avec

$$|S_{ij}^k| = k \text{ tel que : } H(X_i / S_{ij}^k) \text{ est minimum...} \quad (\text{II.4})$$

Dans le cas d'un système déterministe, on a pour chaque variable explicable $X : H(X / \Sigma \setminus X) = 0$, il convient alors de trouver les sous-ensembles explicatifs optimaux pour lesquels l'entropie résiduelle est nulle. On adoptera pour ce faire une démarche descendante basée sur un algorithme exhaustif dont les résultats qui suivent permettent l'accélération.

III.2 Quelques résultats

. Théorème 2421 :

Soit $S \in \mathcal{P}(\Sigma)$

$$\begin{aligned} \forall X, Y \notin S : \text{ si } H(X / S) = 0 \text{ et } H(Y / S \cup X) = 0 \\ \text{ alors : } H(Y / S) = 0 \end{aligned} \quad (\text{II.5})$$

Preuve : -

$$\begin{aligned} H(X/S) = 0 &\Rightarrow H(X \cup S) = H(S) \\ H(Y / X \cup S) = 0 &\Rightarrow H(Y \cup S \cup X) = H(X \cup S) && \Rightarrow \\ \text{or } H(X/S) = 0 &\Rightarrow H(X/S \cup Y) = 0 \Rightarrow H(X \cup S \cup Y) = H(S \cup Y) \end{aligned}$$

$$H(S \cup Y) = H(S) \Rightarrow H(Y/S) = 0$$

Théorème 2422 :

Soit $S \in \mathcal{P}(\Sigma)$

$$\forall X, Y \notin S : \text{si } H(Y/S \cup X) \neq 0 \\ \text{alors : } H(Y/S) \neq 0 \quad (\text{II.6})$$

Preuve : évidente, en effet $S \subset S \cup X$ d'où
 $H(Y/S \cup X) \leq H(Y/S)$

Théorème 2423 :

Soit $S \in \mathcal{P}(\Sigma)$

$$\forall X, Y \notin S : \text{si } H(X/S) \neq 0 \text{ et } H(X/Y \cup S) = 0 \\ \text{alors : } H(Y/S) \neq 0 \quad \forall H(Y/S \cup X) \quad (\text{II.7})$$

Preuve :

$$H(X/S) \neq 0 \Rightarrow H(X \cup S) \neq H(S)$$

$$H(X/S \cup Y) = 0 \Rightarrow H(X \cup Y \cup S) = H(S \cup Y)$$

a) si $H(Y/S \cup X) \neq 0$ alors : $H(Y/S) \neq 0$ (théorème 2422)

b) si $H(Y/S \cup X) = 0$ on a : $H(Y \cup X \cup S) = H(S \cup X)$

on en déduit : $H(X \cup Y \cup S) = H(S \cup X) = H(S \cup Y) \neq H(S)$ d'où le résultat

Remarque 2424 :

Soit $S \in \mathcal{P}(\Sigma)$

$$\forall X, Y \notin S : \text{si } H(X/Y \cup S) \neq 0 \text{ et } H(Y/S \cup X) = 0 \\ \text{alors on ne peut conclure quant à } H(Y/S)$$

En effet :

$$H(X/S \cup Y) \neq 0 \quad \Rightarrow \quad H(X \cup S \cup Y) \neq H(S \cup Y)$$

$$H(X/S) \neq 0 \Rightarrow H(X \cup S) \neq H(S) \quad \Rightarrow \quad H(Y/S) = H(Y/S \cup X) = 0 \Rightarrow H(Y \cup S \cup X) = H(S \cup X)$$

L'ensemble de ces résultats est résumé dans le tableau ci-dessous, dans lequel on note 0 l'état "entropie nulle" et 1 l'état "entropie non nulle".

H(Y/S)		H(X/S ∪ X) - H(X/S)			
		00	01	11	10
H(Y/S ∪ X)	0	0	1	?	
	1	1	1	1	

. La partie hachurée du tableau correspond à une configuration impossible puisque :

$$0 \leq H(X/S \cup Y) \leq H(X/S) \text{ par conséquent :}$$

$$H(X/S \cup Y) > 0 \Rightarrow H(X/S) > 0 \text{ (cf. Théorème 2422)}$$

. L'état de H(Y/S) ne peut être connu dans le cas particulier où :

$$H(Y/X \cup S) = 0, H(X/S \cup Y) \neq 0 \text{ et } H(S/X) \neq 0.$$

Lorsque cette situation se produit, le calcul de H(Y/S) s'avère nécessaire.

D'un point de vue opérationnel, on voit qu'il est possible, dans cinq cas sur six de connaître l'état de H(Y/S), sans aucun calcul, mais seulement à partir de la connaissance des états de H(X/S), H(X/Y ∪ S) et H(Y/X ∪ S). Dans un seul cas sur six, le calcul de H(Y/S) est nécessaire. De tels résultats permettent l'accélération et la mise en œuvre d'un algorithme de modélisation structurale basé sur les deux points suivants :

i) Dans une approche descendante, on part d'abord d'ensembles explicatifs à N éléments, puis on cherche des sous-ensembles explicatifs à N-1, N-2, ... éléments. Lors du calcul de H(X/S), H(Y/S), ... les quantités H(X/S ∪ Y) et H(Y/S ∪ X) sont déjà connues par l'itération précédente.

ii) Lors de la recherche des sous-ensembles explicatifs, pour un système déterministe, il est inutile de continuer à explorer de façon descendante les branches caractérisées par H(Y/S) ≠ 0. En effet, dans ces conditions, la variable Y n'est plus explicable par aucun des sous-ensembles de S.

III.3 Mise en œuvre de l'algorithme de recherche des sous-ensembles explicatifs

Introduction

On considère que le système est déterministe. A partir d'une telle hypothèse, on peut remarquer qu'il suffit qu'une variable X apparaisse dans une relation déterministe pour avoir $H(X/\Sigma X) = 0$. D'une façon générale, pour chaque variable $X_i \in \Sigma$ qui apparaît dans une relation déterministe, on a alors : $H(X_i / \Sigma \setminus X_i) = H(X_i / \theta_i) = 0$. On peut remarquer que cela ne

préjuge en rien du nombre de relations différentes qu'il est possible d'extraire pour décrire le système (nombre de degrés de liberté du système).

Cherchons alors pour chaque variable X_i , $i = 1, \dots, N$ son (ou ses) ensemble explicatif optimal S_i^k tel que :

$$H(X_i / S_i^k) = 0 \text{ et } H(X_i / S_i^k \setminus X_\alpha) > 0 \quad \forall X_\alpha \in S_i^k \quad (\text{II.9})$$

Rappelons que S_i^k peut être unique ou multiple. Par ailleurs, chaque variable X_i , $i = 1, \dots, N$ est considérée comme étant une variable à expliquer et/ou explicative. On cherche dans un premier temps à expliquer chacune des variables du système. La détermination du nombre de degrés de liberté du système interviendra dans un deuxième temps après identification des relations redondantes.

Algorithme I

En tenant compte de l'ensemble des remarques et résultats précédents, celui-ci peut être présenté comme suit :

```

DEBUT
  k ← N-1
  tant que k ≥ 1 faire
    i ← 1
    tant que i ≤ N faire
      j ≠ i
      tant que j ≤ N faire
        si H ( Xi / Sik ) ; Xi ∉ Ski, n'est pas déjà connu , alors le calculer
        connaissant H ( Xj / Sik ∪ Xi ) déduire H ( Xj / Sik )
        si H ( Xi / Sik ) ≠ 0 et H ( Xj / Sik ∪ Xi ) = 0 calculer H ( Xj / Sik )
        j ← j+1
      ft q
    i ← i+1
  ft q
  k ← k-1
ft q
écrire pour chaque Xi, i ∈ { 1, ..., N }
  { Sik } = ⋃j Sjk : l'ensemble de tous les sous-ensembles explicatifs de Xi
FIN
  
```

Optimalité de l'algorithme

Cet algorithme est évidemment optimal étant donné qu'il parcourt, à chaque niveau k , $1 \leq k \leq N$, pour chaque variable, tous les sous-ensembles possibles à la recherche de ceux qui vérifient la condition $H(X_i / S_i^k) = 0$. Son caractère exhaustif garantit l'optimalité des solutions trouvées.

III.4 Les degrés de liberté du système et les sous-ensembles explicatifs multiples d'une variable

L'algorithme descendant précédemment décrit, détermine pour chaque variable X_i , $i \in \{1, \dots, N\}$ ses sous-ensembles explicatifs minimaux. Un sous-ensemble explicatif S_i^k de la variable X_i est qualifié de minimal s'il vérifie la condition suivante :

$$H(X_i / S_i^k) = 0 \text{ et tout } S \in \mathcal{P}(\Sigma) \text{ vérifiant } H(X_i / S) = 0 \text{ est tel que :} \\ \text{card}(S) \geq \text{card}(S_i^k) = k$$

Chaque couple (variable à expliquer, sous-ensemble explicatif) est l'image d'une relation existant dans le système. Le problème qui se pose est celui de fournir toutes les informations nécessaires à l'analyste concernant :

- le nombre du degré de liberté du système ;
- les sous-ensembles explicatifs de chaque variable explicable.

Dans le premier cas, la détermination du nombre de degré de liberté se traduit par la recherche des relations redondantes, alors que dans le deuxième cas, la multiplicité des sous-ensembles fait apparaître une redondance informationnelle.

III.4.1 Degrés de liberté d'un système

Supposons un sous-ensemble de trois variables X_1 , X_2 et X_3 tel que :

$$H(X_1 / X_2 \cup X_3) = H(X_2 / X_1 \cup X_3) = H(X_3 / X_1 \cup X_2) = 0 \quad \text{-- (II.10)}$$

Les trois couples (variable explicable, sous-ensemble explicatif) peuvent être obtenus à partir de l'analyse d'un système, soit :

- à une relation ; par exemple : $a X_1 + b X_2 + c X_3 = 0$ ou bien
- à plusieurs relations ; par exemple :

$$a_1 X_1 + b_1 X_2 + c_1 X_3 = 0, a_2 X_1 + b_2 X_2 + c_2 X_3 = 0$$

Ainsi, par exemple dans le cas d'un système thermodynamique dont le modèle de connaissance fait intervenir entre autres la relation classique $PV/T = R$ où R est une constante, l'application de l'algorithme précédent conduit à :

$$H(P/V/T) = H(V/P/T) = H(T/P/V) = 0$$

On peut déduire d'un tel résultat, trois relations fonctionnelles qui ne sont en réalité que différentes écritures de la seule et unique équation caractéristique des gaz parfaits en état d'équilibre. Ainsi, pour connaître la véritable structure du système, on sera amené à étudier son nombre de degrés de liberté.

Définition : On dira que le système $\Sigma = \{X_1, X_2, \dots, X_N\}$ est de degré $(N - m)$, $1 \leq m \leq N$ si et seulement si il existe au moins une partie de Σ , notée S^m , de cardinal m , telle que :

$$\begin{aligned} H(S^m / \Sigma \setminus S^m) &= 0 \text{ et} \\ H(S^{m+1} / \Sigma \setminus S^{m+1}) &\neq 0 \quad \forall S^{m+1} \in \mathcal{P}_{m+1}(\Sigma) \end{aligned} \quad (\text{II.11})$$

Dans ces conditions, le sous-ensemble de variables S^m est entièrement déterminé par la connaissance des variables $(\Sigma \setminus S^m)$ et aucune partie de Σ à $(m+1)$ éléments ne possède cette propriété.

La relation (II.11) signifie que le plus "grand modèle" que l'on peut extraire de Σ possède $(N-m)$ degrés de liberté. Autrement dit, le plus "grand modèle" permet d'expliquer m variables sachant les valeurs des $(N-m)$ autres.

Détermination des relations redondantes

L'application de l'algorithme précédent conduit à la détermination d'un ensemble de relations entre variables à expliquer et sous-ensembles explicatifs. Certaines relations peuvent se présenter sous différentes écritures, c'est pourquoi, il convient de s'assurer que les relations obtenues ne sont pas redondantes. Nous allons caractériser les cas pour lesquels ceci se produit, ce qui nous permettra de les identifier. Une telle identification conduira à déterminer le nombre des degrés de liberté du système.

Considérons deux variables X_1, X_2 et un sous-ensemble explicatif S tels que :

$$H(X_1 / S \setminus X_1) = H(X_2 / S \setminus X_2) = 0$$

Deux situations peuvent se produire :

- si la connaissance de $S / (X_1 \cup X_2)$ ne permet pas de déterminer uniquement $X_1 \cup X_2$, alors ces deux égalités correspondent à une seule et unique relation.
- si par contre la connaissance de $S / (X_1 \cup X_2)$ conduit effectivement à la détermination de $X_1 \cup X_2$, alors les deux relations sont indépendantes. On peut montrer le théorème suivant :

Théorème II.4.2.2 :

Soit S un sous-ensemble explicatif contenant X_1 et X_2 tel que :

$H(X_1 / S \setminus X_1) = H(X_2 / S \setminus X_2) = 0$, alors
 $H(X_1 / S \setminus (X_1 \cup X_2))$ et $H(X_2 / S \setminus (X_1 \cup X_2))$,
sont tous deux nuls ou tous deux non nuls.

Preuve :

- Supposons $H(X_1 / S \setminus (X_1 \cup X_2)) = 0$ et $H(X_2 / S \setminus (X_1 \cup X_2)) \neq 0$

On a : . $H(X_1 / S \setminus (X_1 \cup X_2)) = H(S / X_2) - H(S / (X_1 \cup X_2)) = 0$
. $H(X_2 / S \setminus X_2) = H(S) - H(S / X_2) = 0$
. $H(S) - H(S / (X_1 \cup X_2)) = 0$

Par ailleurs :

. $H(X_2 / S \setminus (X_1 \cup X_2)) = H(S / X_1) - H(S / (X_1 \cup X_2)) \neq 0$
. $H(X_2 / S \setminus X_1) = H(S) - H(S / X_1) = 0$
. $H(S) - H(S / (X_1 \cup X_2)) \neq 0$

d'où la contradiction

- Supposons maintenant que $H(X_2 / S \setminus (X_1 \cup X_2))$ et

$$H(X_2 / S \setminus (X_1 \cup X_2)) = 0 \Rightarrow H(X_1 \cup X_2 / S \setminus (X_1 \cup X_2)) = 0$$

et le sous-système de variables S est au plus à $|S| - 2$ degrés de liberté. Par ailleurs, si l'un des deux sous-ensembles (S/X_1) ou (S/X_2) est un sous-ensemble explicatif minimal (pour X_1 ou pour X_2) alors :

$H(X_1 / S \setminus X_1) = 0$ et $H(X_2 / S \setminus X_2) = 0$ se réfèrent à une seule relation.

En effet, si $S \setminus X_1$ est sous-ensemble explicatif minimal pour X_1 , on a :

$H(X_1 / S \setminus (X_1 \cup X_2)) \neq 0$ ce qui implique $H(S) - H(S / (X_1 \cup X_2)) \neq 0$ d'où

$$H(X_1 \cup X_2 / S \setminus (X_1 \cup X_2)) \neq 0$$

III.4.2 Les sous-ensembles explicatifs multiples

Dans un système de grande dimension, une même variable peut être liée à d'autres groupes de variables par des relations fonctionnelles différentes. Ceux-ci constituent alors des sous-ensembles explicatifs multiples. De tels sous-ensembles explicatifs multiples de la variable explicable peuvent être de cardinaux identiques ou différents, disjoints ou non. Dans les deux cas, si on note S_1 et S_2 deux sous-ensembles explicatifs de la variable X ,

$X \notin S_1$ et S_2 et $\Sigma = X \cup S_1 \cup S_2$, on montre que :

$I(X : S_1 : S_2) = 2 H(X) + I_X(S_1 : S_2)$, en effet :

$H(X / S_1) = 0 \Rightarrow H(X \cup S_1) = H(S_1)$ et

$H(X / S_2) = 0 \Rightarrow H(X \cup S_2) = H(S_2)$ d'où

$$\begin{aligned} I(X : S_1 : S_2) &= H(X) + H(X \cup S_1) + H(X \cup S_2) - H(X \cup S_1 \cup S_2) \\ &= 2 H(X) + [H(X \cup S_1) - H(X)] - [H(X \cup S_1 \cup S_2) - H(X \cup S_2)] \\ &= 2 H(X) + H(S_1 / X) - H(S_1 / S_2 \cup X) \\ &= 2 H(X) + I_X(S_1 : S_2) \end{aligned}$$

Cette relation s'interprète de la façon suivante :

La quantité d'information circulant entre le triplet (variable explicable X et sous-ensembles explicatifs S_1 et S_2) se divise en deux parties :

- d'une part, l'information apportée en commun par S_1 et S_2 au sujet de X . Celle-ci est redondante, en effet $2 H(X) = I(X : S_1) + I(X : S_2)$: l'un des deux sous-ensembles explicatifs suffit à lui seul à expliquer X , l'autre apparaît comme un sous-ensemble apportant une quantité d'information redondante.
- d'autre part, l'information échangée directement entre S_1 et S_2 indépendamment de X .

Ainsi, la multiplicité des sous-ensembles explicatifs, contient une certaine richesse informationnelle sous forme redondante, qu'il ne serait pas sans intérêt d'exploiter au profit :

- du choix d'un modèle ;
- du diagnostic et/ou la surveillance d'un système.

i) choix et qualité d'un modèle :

A chaque variable explicable X_i est associé un ensemble \mathcal{J}_i des sous ensembles explicatifs multiples S_{ij} . Un tel sous-ensemble s'écrit, si on note k_i son cardinal :

$$\mathcal{J}_i = \{S_{ij}, j = 1, \dots, k_i \text{ tel que : } H(X_i / S_{ij}) = 0\}$$

A chaque sous-ensemble explicatif $S_{ij} \in \mathcal{J}_i$, on associe un coût C_{ij} qu'on appellera : coût d'observation et/ou de mesure d'un tel sous-ensemble. Il est clair qu'un tel coefficient doit faire intervenir essentiellement :

- le coût d'implantation des capteurs (au sens large) nécessaires à l'acquisition des variables explicatives. Ce coût dépend de l'environnement, de la complexité de la procédure de mesure ... etc ;
- le coût d'exploitation du modèle bâti à partir de ces variables : coût énergétique, rapidité d'obtention des valeurs mesurées, fiabilité, maintenabilité, sécurité et disponibilité des capteurs.

Le couple (variable explicable, sous-ensemble explicatif) optimal peut alors être donné par :

$$Q(X_i) = C_i^*(\mathcal{J}_i) = \min_{j=1, \dots, k_i} C_i(S_{ij}), S_{ij} \in \mathcal{J}_i \quad (\text{II.12})$$

où $C_i(S_{ij})$ représente un coût global relatif à l'utilisation du sous-ensemble explicatif S_{ij} pour expliquer la variable X_i . Un tel coût étant à l'évidence sous-additif, la qualité du modèle à (N-m) degrés de liberté retenu pour décrire le système est donnée par :

$$Q \leq \sum_{i=1}^m Q(X_i) = \sum_{i=1}^m C_i^*(\mathcal{J}_i) \quad (\text{II.13})$$

ii) Surveillance du système :

Comme nous l'avons signalé plus haut, l'existence de sous-ensembles explicatifs multiples pour une variable explicable peut faciliter la surveillance ou le contrôle de celle-ci. En effet, si on considère deux ensembles S_1 et S_2 explicatifs de la variable X tels que : $X = f(S_1) = g(S_2)$, l'observation d'une distance entre f et g , nous renseigne sur l'état de fonctionnement des parties de S_1 et S_2 du système. Ainsi, si une défaillance surgit au niveau d'une variable appartenant à S_1 ou S_2 , elle se traduit par une évolution de la distance entre f et g . Une telle défaillance

peut être alors détectée. C'est pourquoi, la redondance d'information véhiculée par les sous-ensembles explicatifs multiples peut être utilisée pour augmenter la fiabilité des systèmes [STAROSWIECKI 1986] [HAMAD 1986].

III.4.3 Remarques

Pour éviter d'augmenter inutilement la dimension du modèle recherché, il convient d'étudier, pour chaque variable explicable, ses sous-ensembles explicatifs optimaux. Nous avons montré, certaines propriétés des ensembles explicatifs qui permettent d'accélérer notablement le processus de leur détermination. Pour des raisons de concision, ce travail a été présenté dans un premier temps dans un cadre déterministe (l'entropie d'une variable modélisée est nulle).

Sous un angle plus général, deux solutions peuvent être envisagées :

1 - On peut interpréter les valeurs de l'indice de modélisabilité en fonction d'un "bruit" supposé superposé à la relation déterministe. Dans ce cas, on peut alors mesurer le risque pris en acceptant ou en rejetant l'hypothèse de dépendance fonctionnelle entre une variable et son ensemble explicatif optimal [SBAI 1983].

2 - On peut proposer une approche de modélisation, dans un cadre non déterministe (présence du bruit sur les modalités du système). C'est ce que nous allons entreprendre dans une deuxième partie de ce chapitre qui est consacrée à la modélisation. En effet, dans un système suffisamment complexe, les données recueillies sur celui-ci sont généralement bruitées, hétérogènes (débit, concentration, P.H., ...) et souvent incomplètes (manque de variables ou modalités explicatives étant donnée qu'en milieu industriel, il est difficile d'accéder à toutes les informations sur le système). Par ailleurs, les systèmes complexes sont, dans la plupart des cas, constitués de sous-systèmes qui interagissent entre eux de façon non simple et régis par des modèles (lorsque ceux-ci existent) non linéaires. Ainsi, le traitement et la validation des données représentant le fonctionnement du système constitue une phase préliminaire nécessaire pour entreprendre la modélisation proprement dite dans des conditions satisfaisantes.

IV - MODELISATION DANS LE CAS GENERAL

IV.1 Position du problème

Dans l'approche précédente, aucune hypothèse n'a été faite initialement au niveau de la nature des variables, chacune d'elles pouvant aussi bien être considérée comme variable à expliquer que variable explicative. Une telle approche est d'une grande utilité, pour l'analyse des systèmes sur lesquels on ne possède aucune information a priori. Dans cette deuxième partie,

nous supposons qu'il a été fixé d'avance une division des variables de Σ en deux groupes.

i) Les variables à expliquer : celles dont l'acquisition est difficile, coûteuse, généralement bruitées, parfois impossibles à mesurer à un instant donné et qu'il convient de prédire ou d'estimer. C'est le cas par exemple d'une station d'épuration des eaux usées où la mesure de la DBO₅ (demande biologique en oxygène) nécessite un délai d'incubation de cinq jours.

ii) Les variables explicatives : ce sont des variables facilement observables dont il est possible de connaître les valeurs à l'aide d'un dispositif de mesure adapté (capteur ou autre). De telles variables sont supposées mesurables, c'est le cas par exemple de la mesure de la concentration en oxygène dissous, des débits, ... etc. dans une station d'épuration des eaux usées. Les variables d'entrée du système devraient faire partie de cette catégorie des variables. Par ailleurs, celles-ci peuvent être soumises à des perturbations de natures diverses et sont susceptibles de fournir des données bruitées, voire erronées.

On note : $Y = \{X_i, i = 1, \dots, p\}$ l'ensemble des variables à expliquer et

$X = \{X_i, i = p + 1, \dots, N\}$ l'ensemble des variables explicatives

$M_X = \{\alpha_i, i = 1, \dots, n\}$

les ensembles de modalités de X et de Y

$M_Y = \{\beta_j, j = 1, \dots, m\}$

Les variables vectorielles X et Y génèrent un tableau de contingence $[P_{ij}]$ dans lequel la case (i,j) contient P_{ij} , probabilité d'occurrence conjointe des modalités α_i et β_j .

On désire analyser les relations entre Y et X à partir des données contenues dans le tableau de contingence. Nous effectuerons cette analyse dans le cas général sans aucune hypothèse particulière sur le modèle du système.

IV.2 Approche préliminaire

La modélisation ou l'explication d'un système nécessite la décomposition de ses variables en deux groupes : les variables à expliquer notées Y et les variables explicatives notées X. L'intensité de liaison entre ces deux variables peut être mesurée à l'aide de l'entropie conditionnelle $H(Y/X)$ qui renseigne sur la possibilité d'explication de Y par X. En effet, trois cas peuvent être distingués :

IV.2.1 - $H(Y/X) = H(Y)$:

Indépendance statistique entre Y et X, ceci arrive rarement dans la réalité, puisque s'il on prétend expliquer Y par X, il convient de choisir un ensemble explicatif apportant le maximum d'information à son sujet.

IV.2.2 - $H(Y/X) = 0$:

Liaison totale entre Y et X (système déterministe : on se retrouve dans le cas précédent) et $I(Y : X) = H(X)$. Il est alors possible de proposer un modèle à caractère déterministe puisque :

$$H(Y/X) = 0 = \sum_i P(X=\alpha_i) H(Y/X=\alpha_i) \Rightarrow$$

$$H(Y/X=\alpha_i) = - \sum_j P(Y=\beta_j / X=\alpha_i) \text{Log} P(Y=\beta_j / X=\alpha_i) = 0 \quad \forall \alpha_i \in M_X$$

étant donné que $P(X = \alpha_i)$ est toujours ≥ 0 . Or la fonction $z \log z$ ne s'annule que pour $z = 0$ et $z = 1$. Cette dernière égalité implique que pour chaque i il existe un seul j tel que : $p(Y = \beta_j / X = \alpha_i) = P_{j/i} = 1$; celui-ci est forcément unique puisque

$$\sum_j P_{j/i} = 1$$

Autrement dit, il existe une relation univoque entre Y et X, c'est-à-dire :

$$\exists \text{ une fonction } F : M_X \rightarrow M_Y \text{ telle que : } Y = F(X) \Leftrightarrow P_X \leq P_Y$$

où " \leq " signifie que la partition de X est plus fine que ou égale à la partition de Y (cf relation I.7 chapitre I). Dans ce cas, on peut associer à chaque modalité de Y une ou plusieurs modalités de X, on peut alors établir un modèle à p équations parmi les N variables, donc un système à $(N-p)$ degrés de liberté.

IV.2.3 - $0 < H(Y/X) < H(Y)$

Y n'est expliqué que partiellement par X, cette absence d'explication totale peut être due à des causes d'origines diverses que l'on peut diviser en deux groupes :

i) absence d'explication à tort :

Le modèle déterministe $Y = f(X)$ est représentatif du système, cependant les données relevées sur celui-ci sont biaisées :

- . la présence d'un bruit important : la mesure d'une grandeur trop bruitée ou son acquisition par un dispositif de mesure inadapté peut conduire à des données erronées ;
- . le mauvais fonctionnement d'un sous-système du système global peut être à l'origine également de mesures erronées.

ii) absence d'explication à raison :

- . quelques variables pertinentes, sous estimées, ignorées ou laissées volontairement de côté conduisent à une description incomplète du système. Cette situation s'explique par le fait qu'il y a trop de variables à expliquer et/ou trop peu de variables explicatives ;
- . une variable explicative peut fournir une quantité d'information sous forme redondante et ne peut améliorer l'explication de la variable à expliquer ;
- . le système est suffisamment complexe et ne peut être appréhendé par un modèle déterministe.

Par ailleurs lorsque $H(Y/X) > 0$, on peut montrer facilement qu'il existe au moins une variable $Y_j \in Y$ telle que :

$$H(Y_j/X) > 0$$

En effet, d'après la propriété de sous-additivité de l'entropie (cf I.28) et d'après le principe de conditionnement uniforme (cf I.65), on peut alors écrire :

- . $H(X_1) + H(X_2) + \dots + H(X_j) + \dots + H(X_p) \geq H(Y) > 0$
- . $H(X_1/X) + H(X_2/X) + \dots + H(X_j/X) + \dots + H(X_p/X) \geq H(Y/X) > 0$,

par conséquent $H(Y/X) > 0$ signifie qu'il existe au moins une variable $X_j \in Y$ telle que son entropie est non nulle même si, dans le cas limite où l'entropie des autres variables de Y sont toutes nulles. Cette absence d'explication peut être causée (à tort) par la présence du bruit sur les modalités de Y (problème d'acquisition) ou causée (à raison) par l'insuffisance des variables explicatives (problème de modélisation). Il convient alors d'identifier les modalités de Y réellement mal expliquées et dont l'explication requiert l'introduction des variables explicatives supplémentaires. Mais avant de proposer une approche permettant l'identification de telles modalités, nous allons rappeler les approches classiques et nous montrerons leurs limites en présence du bruit sur les variables du système.

IV.3 Les approches classiques

On peut distinguer deux catégories de méthodes de modélisation structurale que l'on peut résumer comme suit :

- . les approches générales (CONANT, RICHETIN, TORO et STAROSWIECKI) ;
- . les approches utilisant un modèle (RINK, DUFOUR, PRADIN, MARFAING).

IV.3.1 Les approches générales

Celles-ci ne supposent aucun modèle particulier, et visent à trouver les relations essentielles de causalité entre variables du système à partir des données relevées sur celui-ci. Ces méthodes sont basées généralement sur le calcul d'un indice de couplage entre variables du système qui peut être aussi bien statique que dynamique. Certains de ces indices sont discutés au chapitre I § IV.2.6. Dans ce sens, pour juger la faisabilité du problème de modélisation, TORO [1982] propose l'utilisation des indices suivants :

$$\bullet \quad \% H(X_i / \theta_i) = \frac{H(X_i / \theta_i)}{H(X_i)} \quad \text{avec} \quad \theta_i = \Sigma \setminus X_i \quad (\text{II.14})$$

qui représente le pourcentage non explicable ou non modélisable de l'entropie de X_i . Le complément à 1 de cet indice mesure le pourcentage explicable ou modélisable de l'entropie de X_i .

$$\bullet \quad \% \text{TNE}(Y/X) = \frac{\sum_{i=1}^P H(X_i / \theta_i)}{\sum_{i=1}^P H(X_i)} \quad (\text{II.15})$$

qui représente le pourcentage du total non explicable dans le cas où X est l'ensemble des variables explicatives et Y l'ensemble des variables à expliquer et/ou explicatives. Ainsi une valeur proche de zéro de cet indice indique qu'il y a suffisamment de redondance dans le système ou bien que les relations de causalité entre variables à expliquer et variables explicatives sont fortes. Par contre, une valeur proche de 1 de cet indice montre qu'il y a trop de variables à expliquer et/ou trop peu de variables explicatives et qu'il y a très peu de chances d'établir un modèle du système. En effet, la possibilité de construire un tel modèle dépend de la mesure dans laquelle le système "s'explique en lui-même", autrement dit, dans quelle mesure on n'est pas obligé d'aller chercher des explications dans des variables ou objets externes à Σ .

IV.3.2 Les approches utilisant un modèle

Ces approches sont généralement basées sur l'utilisation d'un modèle analytique linéaire liant les variables d'entrée X aux variables de sortie Y au moyen de relations de récurrence ou d'état. Dans ce sens, J. DUFOUR [1979] propose le modèle récurrent suivant :

$$Y(k) = \sum_{i=0}^q a_i X(k-i) \quad (\text{II.16})$$

ou bien

$$Y(k) = \sum_{i=1}^q a_i Y(k-i) + b_1 X(k) \quad (\text{II.17})$$

utilisant les concepts d'entropie et de transinformation pour la caractérisation de ces types de systèmes : détermination de l'ordre q du système et du décalage temporel r dans le cas des systèmes dynamiques ; coefficients qui interviennent dans des expressions du type :

$$I \{ Y(k) : Y(k+1) : \dots : Y(k+q) : \dots : Y(k+q+r) \} \quad (\text{II.18})$$

Une étude originale a conduit DUFOR à proposer dans le cas des entrées gaussiennes, des relations entre transmutations qui autorisent la caractérisation et parfois l'identification des systèmes linéaires. S'inspirant des travaux de RINK [1973], il indique qu'il est possible de traiter certains systèmes non linéaires d'autant plus que les concepts informationnels s'appliquent aussi bien à des relations linéaires que non linéaires.

Toujours sous l'hypothèse d'un modèle linéaire, [MARFAING 1983] [PRADIN 1986] supposent que les données à analyser sont issues d'un système dynamique dont l'évolution est décrite par :

$$X(k+1) = A X(k) + B U(k) \quad (\text{II.19})$$

$$Y(k) = C \cdot X(k)$$

où $X(k)$ est le vecteur d'état de dimension N , $U(k)$ le vecteur de commande et $Y(k)$ le vecteur de sortie de dimension $M < N$. PRADIN propose de chercher la structure de la matrice C , calquée sur celle de la matrice A . Pour mettre en évidence aussi bien les couplages statiques que dynamiques dus aux relations entre les différentes variables de sortie et d'état, l'auteur définit un indice de dissimilarité dynamique noté d'_{ij} et s'écrit :

$$d'_{ij}(Y_i : Y'_j) = 1 - \frac{I(Y_i : Y_j \cup Y'_j)}{H(Y_i \cup Y_j \cup Y'_j)} \quad (\text{II.20})$$

Si la théorie de l'information apporte une aide importante par la possibilité de définir, à partir des données, des indices de dissimilarité entre variables et groupes de variables entraînant une réduction importante d'information à traiter [PRADIN 1986], elle fournit par contre des indices de dissimilarité dont il faut connaître les limites dans le cas dynamique. En effet, comme le

signale l'auteur, ces indices permettent difficilement la distinction entre les couplages structurels et les couplages apparents qui peuvent apparaître sous l'effet de plusieurs facteurs.

Remarque :

Toutes les approches précédentes sont globales. Elles analysent l'ensemble des échantillons caractérisant le système et de ce fait, elles aboutissent à des résultats qui traduisent le comportement moyen de ces échantillons. Le tableau de contingence ainsi généré n'étant pas filtré, ces méthodes ne peuvent distinguer entre variables non expliquées à tort ou à raison. C'est pourquoi, nous proposons d'effectuer des analyses locales, ce qui nous permettra de proposer un algorithme de filtrage du tableau de contingence.

IV.4 L'approche proposée [BARBOUCHA 1987]

Nous allons montrer dans un premier temps l'insuffisance des approches classiques de modélisation, lorsque le bruit vient se superposer aux modalités du système. Pour des systèmes suffisamment complexes, l'expérience prouve qu'il est difficile d'expliquer une variable de façon satisfaisante sur l'ensemble de ses modalités. Dans ce sens, nous proposons des tests permettant la détection des modalités bien expliquées et des modalités mal expliquées et dont l'explication nécessite l'introduction des variables explicatives supplémentaires. A cet effet, nous proposons un algorithme de recherche de ces sous-ensembles explicatifs supplémentaires.

IV.4.1 Insuffisance des approches classiques

Toutes les approches classiques sont basées sur le calcul de l'entropie conditionnelle ou la transinformation à partir des données brutes recueillies sur le système. Or, comme nous l'avons signalé au chapitre I, § 4.2.7.2., le calcul de tels indices sur les données observées risque d'être insuffisant pour conclure sur la faisabilité du problème de modélisation. En effet, si on reprend les hypothèses formulées au § 4.2.7.2 :

. H_0 : liaison fonctionnelle bruitée entre Y et X, on a :

$$H(Y/X)_0 = H(Y/X = \alpha_i)_0 = -P_{j/i} \text{Log } P_{j/i} - (1 - P_{j/i}) \text{Log } \frac{1 - P_{j/i}}{m - 1}$$

$$I(Y : X)_0 = \text{Log } m + P_{j/i} \text{Log } P_{j/i} + (1 - P_{j/i}) \text{Log } \frac{1 - P_{j/i}}{m - 1}$$

. H_1 : liaison fonctionnelle partielle non bruitée entre Y et X, on a :

$$H(Y/X)_1 = H(Y/X = \alpha_i)_1 = -P_{j/i} \text{Log } P_{j/i} - (1 - P_{j/i}) \text{Log } (1 - P_{j/i})$$

$$I(Y : X)_1 = \text{Log } m + P_{j/i} \text{Log } P_{j/i} + (1 - P_{j/i}) \text{Log } (1 - P_{j/i})$$

Le calcul de la différence entropique et transinformationnelle donne :

$$I(Y : X)_1 - I(Y : X)_0 = \Delta I = H(X/X)_0 - H(Y/X)_1 = \Delta H$$

$$\text{avec } \Delta H = \Delta I = (1 - P_{ji}) \text{Log}(m - 1) \quad (\text{II.21})$$

Une telle relation indique que l'entropie conditionnelle sous H_0 qui correspond à une relation fonctionnelle bruitée est plus élevée que l'entropie conditionnelle sous H_1 où X ne peut expliquer à lui seul Y et particulièrement si $(1 - P_{ji})$ est non négligeable et représente un phénomène réel. Par conséquent, la présence d'un bruit de cette nature sur les modalités de Y (hypothèse H_0) conduit à une valeur élevée de $H(Y/X)$ et de ce fait, une conclusion hâtive affirmerait soit l'impossibilité d'établir un modèle déterministe ou une description insuffisante du système.

Autrement dit, une approche de modélisation bâtie uniquement sur les indices informationnels peut être mise en défaut si les mesures effectuées sur le système présentent un bruit non négligeable. Une telle approche peut s'avérer incomplète et ne peut distinguer d'une manière opérationnelle entre les modalités de Y non expliquées à tort (à cause de la présence du bruit) ou non expliquées à raison (manque de variables explicatives). D'où la nécessité de faire appel à une approche permettant la détection de telles modalités.

IV.4.2 Test de détection des modalités mal expliquées

Il est clair que la résolution d'un tel problème nécessite un moyen permettant de tester les deux hypothèses suivantes :

- h_0 : Y est non expliquée à tort à cause de la présence du bruit sur ses modalités ;
- h_1 : Y est non expliquée à raison à cause essentiellement d'un manque de variables explicatives ou une description insuffisante du système.

Pour distinguer ces deux hypothèses, nous ferons appel dans un premier temps aux notions de tests statistiques et nous proposerons ensuite une approche basée sur l'étude de la répartition des résidus et de la nature de la quantité d'information qu'ils recèlent.

IV.4.2.1 Les tests statistiques

Pour la mise en œuvre de ces tests, il convient de fixer un modèle de comportement du bruit. En adoptant un modèle de comportement uniforme, la distribution de probabilité de chacune des lignes i , $i \in \{1, \dots, n\}$ du tableau de contingence s'écrit

en notant $P_{j/i} = \max_k P_{k/i}$, $k \in \{1, \dots, m\}$:

$$\tilde{P}(Y/X=\alpha_i) = \left\{ P_{j/i}, \frac{1 - P_{j/i}}{m-1}, \dots, \frac{1 - P_{j/i}}{m-1} \right\} \quad (\text{II.22})$$

A cette distribution, représentant le modèle de comportement d'une modalité de Y non expliquée à tort, est comparée la distribution réelle $P(Y/X = \alpha_i)$ dont les éléments sont ordonnés, c'est-à-dire :

$$P(Y/X = \alpha_i) = \{P_{j/i}, j = 1, \dots, m \text{ tels que : } P_{1/i} \geq \dots \geq P_{m/i}\} \quad (\text{II.23})$$

La comparaison de ces deux distributions se fait, d'une manière classique, par la mise en œuvre des tests statistiques [BENZECRI 1980]. Le test du χ^2 est sans conteste le plus utilisé. Appliqué, aux deux lois, celui-ci a pour but de vérifier dans quelle mesure la distribution réelle mesurée P et la distribution modèle \tilde{P} supposée le résultat de la superposition d'un phénomène réel et d'un phénomène purement aléatoire de distribution uniforme, décrivent de manière acceptable le même sous-échantillon $L\alpha_i$ de L. Celui-ci s'écrit :

$$\chi^2 = L\alpha_i = \sum_{j=2}^m \frac{\left(P_{j/i} - \frac{1 - P_{j/i}}{m-1} \right)^2}{\left(\frac{1 - P_{j/i}}{m-1} \right)} \quad (\text{II.24})$$

A partir d'un seuil ϵ , du χ^2 calculé et du χ^2 tabulé, il est alors possible d'accepter ou de rejeter l'hypothèse h_0 . Par ailleurs, on peut remarquer que lorsque la distribution mesurée se réduit à un seul élément de probabilité égale à 1, ou bien lorsqu'elle est identique à la distribution du modèle, la valeur du χ^2 sera égale à zéro. Par contre, dans le cas où elle est formée de deux éléments de probabilité $P_{j/i}$ et $(1 - P_{j/i})$, le χ^2 sera égal à :

$$\chi^2 = L \alpha_i (1 - P_{j/i}) \cdot (m - 2) \quad (\text{II.25})$$

Ce test qui semble bien convenir pour la comparaison des deux lois P et \tilde{P} possède toutefois un inconvénient dans le cas particulier où la ligne étudiée est totalement bruitée et ne contient aucun phénomène réel. En effet, dans cette situation qui arrive cependant rarement dans la pratique, le test conduit à une faible valeur du χ^2 , l'hypothèse h_0 sera retenue, alors que l'on se trouve manifestement dans l'hypothèse h_1 . Pour pallier à cet handicap, nous allons proposer une approche qui consiste à étudier la variation entropique au niveau de chaque ligne du tableau de contingence.

IV.4.2.2 Variation entropique

Ordonnons les éléments de la distribution $P (Y/X = \alpha_i)$ par ordre décroissant et appelons Y_2 la variable Y après réunion de ses deux premières modalités. Moyennant une renumérotation donnée, le calcul de l'entropie de Y avant et après un tel regroupement donne :

$$\bullet H (Y/X=\alpha_i) = - P_{1/i} \text{Log } P_{1/i} - \sum_2^m P_{j/i} \text{Log } P_{j/i} \quad (\text{II.26})$$

$$\bullet H (Y_2 / X=\alpha_i) = - (P_{1/i} + P_{2/i}) \text{Log } (P_{1/i} + P_{2/i}) - \sum_3^m P_{j/i} \text{Log } P_{j/i} \quad (\text{II.27})$$

En notant $P_{2/i} = \delta P_{1/i}$, avec $0 \leq \delta \leq 1$, la variation entropique

$$\Delta H = H (Y/X = \alpha_i) - H (Y_2/X = \alpha_i)$$

provoqué par un tel regroupement est proportionnelle à la réduction d'incertitude sur Y et s'écrit :

$$\Delta H = P_{1/i} \cdot [(1 + \delta) \text{Log } (1 + \delta) - \delta \text{Log } \delta] \quad (\text{II.28})$$

résultat qui s'interprète comme la quantité d'information perdue à l'issue de l'opération de regroupement des deux premières modalités de Y . Une telle fonction peut s'écrire sous forme normalisée (coefficient compris entre 0 et 1) :

$$\eta = \frac{\Delta H}{2 P_{1/i}} = \frac{(1 + \delta) \text{Log } (1 + \delta) - \delta \text{Log } \delta}{2} \quad (\text{II.29})$$

. Lorsque δ est au voisinage de zéro alors η tend vers zéro puisque $P_{2/i}$ est trop faible devant $P_{1/i}$ et correspond à la probabilité d'une modalité due à la présence du bruit. On peut alors considérer, dans ce cas, que Y est non expliqué à tort à cause de la présence du bruit sur ses modalités (hypothèses h_0).

. Lorsque δ est non négligeable et particulièrement lorsqu'elle tend vers 1, alors $\eta \rightarrow 1$ puisque $P_{2/i}$ est du même ordre de grandeur que $P_{1/i}$. Or ceci, peut correspondre à deux situations différentes :

- Soit $P_{1/i}$ est élevé donc $P_{2/i}$ l'est également et proviennent tous les deux de l'existence de deux phénomènes réels et on se trouve alors dans le cas où X ne peut expliquer à lui seul ces deux modalités de Y (hypothèse h_1) :

$$P_{2/i} \rightarrow P_{1/i} \Rightarrow \delta \rightarrow 1 \Rightarrow \eta \rightarrow 1$$

- Soit $P_{1/i}$ est faible, il en est de même pour $P_{2/i}$; ceci correspond au cas où une ligne est totalement bruitée. Le test doit, confirmer que le système est dans l'hypothèse h_1 (absence de variables réellement explicatives). En effet :

$$P_{2/i} \rightarrow P_{1/i} \Rightarrow \delta \rightarrow 1 \Rightarrow \eta \rightarrow 1 \quad \text{d'où l'hypothèse } h_1$$

Le choix d'un seuil de variation entropique limite ε , permet de distinguer nettement les deux hypothèses retenues h_0 et h_1 .

Nous traçons ci-dessous (fig. 1) la courbe $\eta = f(\delta)$

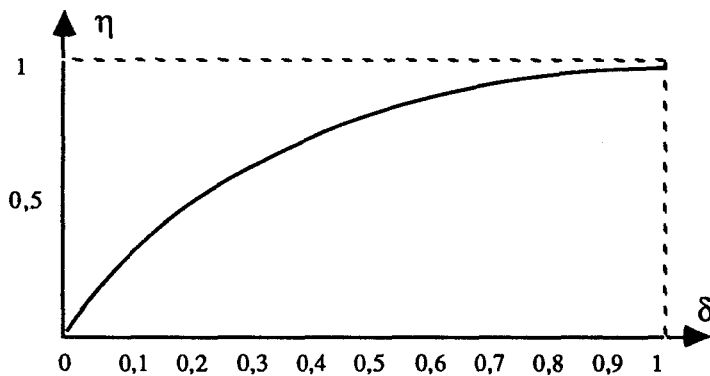


Fig I.1 : Tracé de $\eta = f(\delta)$

En définitive, à l'issue de cette étape, il est possible de décomposer l'ensemble des modalités de X en deux sous-ensembles M_X^1 et M_X^2 tels que :

$$\cdot M_X^1 = \{\alpha_i, i = 1, \dots, v \text{ tels que : } P(Y/X = \alpha_i) \text{ vérifie } h_0\} \quad (\text{II.30})$$

$$\cdot M_X^2 = \{\alpha_i, i = v + 1, \dots, n \text{ tels que : } P(Y/X = \alpha_i) \text{ vérifie } h_1\}$$

On a ainsi réalisé une partition de l'ensemble des modalités de X en un sous-ensemble de modalités explicatives M_X^1 et un sous-ensemble de modalités M_X^2 qui ne peuvent expliquer à elles seules la partie non encore expliquée (ou modélisée) de Y . Le recours à des variables explicatives supplémentaires est donc nécessaire pour l'explication totale de Y .

IV.4.3 Recherche des variables explicatives supplémentaires

La modélisation des procédés industriels complexes passe nécessairement par une phase de saisie d'informations représentant l'état de ces procédés. Lorsque ceux-ci sont suffi-

samment complexes et de grande dimension, il est difficile de tenir compte dans la phase de modélisation de l'ensemble des variables et paramètres caractérisant leurs fonctionnements. Ainsi, lorsqu'on procède à la sélection des variables en vue de décrire le comportement d'un système, on s'aperçoit parfois que la description est incomplète à cause de l'insuffisance des variables explicatives. La description complète du système nécessite de créer de manière itérative des problèmes d'explication successifs, pour lesquels il s'agit de définir des variables explicatives supplémentaires optimales. Cette recherche peut s'effectuer soit parmi les variables secondaires (disponibles ou candidates), mais peut aussi se traduire par la création de nouvelles variables (synthèse structurale : cas des systèmes logiques par exemple).

IV.4.4 Algorithmes

IV.4.4.1 Algorithme 2 : cas des variables secondaires disponibles

Comme nous l'avons dit précédemment, lors de la première phase de modélisation d'un système de grande dimension, on sélectionne certaines variables explicatives (variables primaires) et on laisse volontairement de côté quelques variables (variables secondaires ou candidates) pour diverses raisons : encombrement, coût du dispositif de mesure, augmentation de la dimension du modèle ..., etc. En effet, ceci arrive assez fréquemment lorsque le système possède des modes de fonctionnement multiples, chacun d'eux étant régi par un ensemble d'équations qui le caractérisent. Autrement dit, une approche locale peut suffire à appréhender le système dans une zone de fonctionnement donnée, alors que son explication totale nécessite une approche globale faisant intervenir l'ensemble des variables explicatives aussi bien primaires que secondaires candidates. Dans ce dernier cas, si on suppose que les relevés de ces variables candidates sont disponibles, on peut alors proposer un algorithme permettant de les sélectionner de façon optimale.

Si on note :

$Z = \{X_i, i = N + 1, \dots, M\}$: l'ensemble des variables secondaires candidates

et

$M_Z = \{\gamma_i, i = 1, \dots, r\}$: l'ensemble de leurs modalités respectives

le génération des sous-ensembles explicatifs optimal $S^* \in \overline{\mathcal{P}(Z)}$, avec $\mathcal{P}(Z)$: l'ensemble des parties de Z , s'effectue en trois étapes successives.

i) Pour chaque $\alpha_i \in M_X^2, i \in \{v + 1, \dots, n\}$, on définit l'ensemble \mathcal{J}_i des sous-ensembles explicatifs de $Y/X = \alpha_i$. Un tel sous-ensemble S est tel que :

$$\forall S \in \mathcal{J}_i \quad \Delta H (Y / (X = \alpha_i) \cup (S = \gamma)) \leq \varepsilon \quad (\text{II.31})$$

où ϵ est un seuil de variation entropique limite que l'on se fixe.

ii) L'ensemble explicatif optimal est alors défini de la façon suivante :

$$\mathcal{Y} = \left\{ S, \text{ tel que : } S = \bigcup_{i \in I} S_i \quad \text{avec } S_i \in \mathcal{Y}_i \quad \text{et } I = (v+1, \dots, n) \right\}$$

L'ensemble de tous les sous-ensembles explicatifs possibles de Y

Sachant $X = \alpha, \forall \alpha \in M_X^2$

iii) Parmi tous ces sous-ensembles, nous choisissons celui de cardinal minimum noté S^* :

$$S^* \in \mathcal{Y} \quad \text{tel que : } \text{card}(S^*) \leq \text{card}(S); \forall S \in \mathcal{Y}$$

L'ordre de cet algorithme noté OA est :

$$OA \leq (n - v) \cdot (2^{M-N} - 1) \quad (\text{II.32})$$

puisque l'examen des sous-ensembles explicatifs de chaque modalité nécessite 2^{M-N} opérations.

Cet algorithme est évidemment optimal puisque S^* est de cardinal minimum et permet conjointement avec X l'explication totale de Y. Il peut être également unique ou multiple. Dans ce dernier cas, une étude qualitative du modèle peut alors être opérée pour le choix de celui minimisant un critère de coût (cf II.12).

D'une manière opérationnelle, la recherche des sous-ensembles explicatifs pour chaque modalité $\alpha_i, i = v+1, \dots, n$, peut se faire à l'aide :

☞ d'un algorithme ascendant (AA) qui élabore de façon itérative une chaîne de sous-ensembles explicatifs tels que :

$$- \quad X \cup S^{k+1} = (X \cup S^k) \cup \{X_\alpha\}, \forall X_\alpha \in Z \quad (\text{II.33})$$

X_α étant choisie de façon optimale parmi les variables candidates non encore utilisées.

☞ d'un algorithme descendant (AD) utilisant la procédure inverse. Opérant entre les niveaux k et k-1 de $\mathcal{P}(Z)$, celui-ci construit une chaîne de sous-ensembles tels que :

$$- \quad X \cup S^{k-1} = (X \cup S^k) \setminus \{X_\alpha\}, \forall X_\alpha \in \mathcal{P}(Z) \quad (\text{II.34})$$

La combinaison des opérateurs AD et AA [TORO 1982] conduit à un algorithme de transfert opérant dans les niveaux $k-1$, k et $k+1$. Le caractère contractant des opérateurs AAoAD et ADoAA implique la convergence de l'algorithme vers un sous-ensemble explicatif optimal.

IV.4.4.2 Synthèse structurale

Dans ce cas, on dispose à l'issue du test précédent d'un certain nombre de modalités de X non explicatives, égal à $n-v$. Autrement dit, à chaque modalité $\alpha_i \in M_X^2$ sont associées au moins deux modalités différentes de Y. L'explication totale de cette variable nécessite la création de modalités séparatrices permettant de distinguer les modalités non expliquées par X. Pour créer ces nouvelles modalités, les deux approches suivantes peuvent être utilisées séparément ou en combinaison.

1) Faire appel à des variables supplémentaires :

On peut les choisir parmi un ensemble de variables supplémentaires disponibles (c'est le cas précédent : ALGORITHME 2) ; elles correspondent alors à des grandeurs existant réellement au niveau du processus et ont une interprétation physique. Si cela n'est pas possible, on peut introduire une variable explicative à nombre de modalités minimal, notée Z, telle que : $H(Y/X \cup Z) = 0$. Par rapport au système modélisé, cette variable n'a pas d'interprétation physique précise. On peut éventuellement l'interpréter en terme d'état, et/ou utiliser l'approche ci-dessous.

2) Utiliser les valeurs passées des variables disponibles :

Cela revient à supposer que l'ordre (la dimension p de l'horizon d'observation) retenu initialement pour décrire le système est trop faible et ne peut générer un nombre d'états suffisants pour expliquer Y totalement. L'augmentation de l'ordre du système revient à tenter d'expliquer la variable Z par l'intermédiaire des valeurs passées des variables explicatives. Deux applications immédiates de ces approches sont les suivantes :

i) les systèmes inversibles :

Lors de l'analyse d'un système, on est amené parfois à chercher des relations biunivoques entre variables à expliquer et variables explicatives. Selon la nature du système analysé, on peut se trouver dans les deux situations suivantes :

- le système peut être inversible en utilisant uniquement les variables retenues pour sa description. Dans ce cas, il est clair que X et Y sont tels que $H(Y/X) = 0$ et $H(X/Y) = 0$.
- le système est initialement non inversible, il est alors nécessaire de faire appel à des

variables explicatives supplémentaires. Le théorème ci-dessous précise une telle situation.

Théorème :

si $H(Y/X) = 0$ et s'il est possible de trouver $Z \in \mathcal{P}(Z)$ tel que :
 $H(X/Y \cup Z) = 0$ avec $H(Z/X) = 0$
 alors le système $(X, Y \cup Z)$ est totalement inversible

Preuve :

$$H(Y/X) = 0 \Rightarrow H(Y/X \cup Z) = 0$$

$$H(Y \cup Z/X) = 0$$

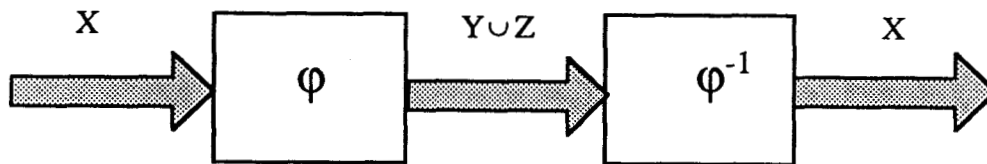
$$H(Z/X) = 0 \Rightarrow H(Z/X \cup Y) = 0$$

1) or $H(Y \cup Z/X) = 0 \Leftrightarrow P_X \leq P_{Y \cup Z} \Rightarrow \exists \varphi : M_X \rightarrow M_{Y \cup Z}$ tel que : $Y \cup Z = \varphi \circ X$

2) $H(X/Y \cup Z) = 0 \Leftrightarrow P_{Y \cup Z} \leq P_X \Rightarrow \exists G : M_{Y \cup Z} \rightarrow M_X$ tel que : $X = G \circ (Y \cup Z)$

$$(1) \text{ et } (2) \Rightarrow P_X = P_{Y \cup Z}$$

d'où d'après (I.6 chapitre I) : $G = \varphi^{-1}$ et $\varphi \circ \varphi^{-1} = I$ (application identité) ; par conséquent est une application biunivoque.



ii) modélisation des systèmes séquentiels : synthèse des variables secondaires

Considérons une machine séquentielle décrivant le modèle d'un système donné. Celle-ci est généralement définie par le quintuplet :

$$M = (X, Y, Q, \delta, \lambda) \text{ où :}$$

X est la variable d'entrée à n modalités

Y est la variable de sortie à m modalités

Q est l'ensemble fini des états internes à la modalité

$\delta : Q \times X \rightarrow Q$ est la fonction état suivant ou de transition

$\lambda : Q \times X \rightarrow Y$ pour une machine de MEALY

: $Q \rightarrow Y$ pour une machine de MOORE constitue la fonction de sortie.

La détermination de la fonction de sortie Y nécessite la synthèse d'une nouvelle variable secondaire Z permettant de distinguer les différents états Q de la machine séquentielle possédant les mêmes entrées et des sorties différentes. Ces variables sont appelées également variables internes et permettent une mémorisation des états de la machine. Autrement dit, la modélisation d'un système séquentiel nécessite la synthèse de variables explicatives supplémentaires capables de mémoriser les évolutions passées et caractériser tout changement ou évolution de son état interne. Cette représentation implique le choix d'un codage respectant les adjacences des états. Il existe de nombreuses méthodes de codage, les critères guidant leur mise en œuvre peuvent être par exemple, en plus du critère précédent [COURVOISIER 1986] :

- l'obtention d'un circuit de complexité minimale ;
- l'obtention d'un circuit à temps de réponse minimal ... etc.

Ceci revient en réalité, à générer un ensemble explicatif optimal minimisant une fonction coût caractérisant sa qualité [cf relation II.13].

Une application immédiate de la synthèse structurale est par exemple la recherche de l'inverse d'un circuit combinatoire. En effet, si on note X l'entrée d'un tel circuit et Y sa sortie, sachant que $H(Y/X) = 0$, le calcul de $H(X/Y)$ permet de conclure sur son inversibilité. Ainsi si $H(X/Y) = 0$, alors le circuit est inversible, dans le cas contraire, la concaténation des variables de sortie et d'une variable secondaire Z est nécessaire. Son codage dépend du nombre de modalités de X non expliquées par Y .

EXEMPLE :

Soit le système combinatoire donné par le tableau ci-dessous où X et Y contiennent respectivement trois variables X_1, X_2, X_3 et Y_1, Y_2, Y_3 .

X			Y			Z	
X_1	X_2	X_3	Y_1	Y_2	Y_3	Z_1	Z_2
0	0	0	0	0	0	∅	∅
0	0	1	0	0	1	∅	∅
0	1	0	0	1	0	∅	∅
0	1	1	1	1	1	0	0
1	0	0	1	0	0	∅	∅
1	0	1	1	1	1	0	1
1	1	0	1	1	1	1	1
1	1	1	1	1	1	1	0

$H(Y/X) = 0$, en effet, ceci correspond à la fonction logique suivante :

$$Y_1 = X_1 + X_2 X_3$$

$$Y_2 = X_2 + X_1 X_3$$

$$Y_3 = X_3 + X_1 X_2$$

. $H(X/Y) > 0$, la recherche de l'inverse nécessite l'introduction d'une variable Z telle que $H(X/YUZ) = 0$. Une telle variable doit faire apparaître 4 modalités séparatrices. Compte tenu que les variables du système sont binaires, Z doit contenir deux variables Z_1 et Z_2 .

Les modalités séparatrices sont représentées sur le tableau. La notation ϕ indique une valeur indifférente. Les fonctions ci-dessous donnent une réalisation possible du système inversible.

$$\left\{ \begin{array}{l} X_1 = Y_1 \bar{Y}_2 \bar{Y}_3 + Y_1 Y_2 Y_3 Z_2 + Y_1 Y_2 Y_3 Z_1 \\ X_2 = \bar{Y}_1 Y_2 \bar{Y}_3 + Y_1 Y_2 Y_3 Z_1 + Y_1 Y_2 Y_3 \bar{Z}_2 \\ X_3 = \bar{Y}_1 \bar{Y}_2 Y_3 + Y_1 Y_2 Y_3 \bar{Z}_1 + Y_1 Y_2 Y_3 \bar{Z}_2 \end{array} \right. \quad \text{avec} \quad \left\{ \begin{array}{l} Z_1 = X_1 X_2 \\ Z_2 = \bar{X}_2 + \bar{X}_3 \end{array} \right.$$

La recherche de l'inverse d'un circuit combinatoire permet d'envisager des structures auto-déTECTABLES pour la résolution des déTECTIONS en ligne des pannes éventuelles par des procédures matérielles [BERTRAND 1974] [MERCIER 1977].

V - APPLICATION : MODELISATION D'UNE STATION D'EPURATION DES EAUX USEES UTILISANT LA VOIE BIOLOGIQUE

V.1 Description d'une station d'épuration [BARBOUCHA 1981]

Une station d'épuration est généralement composée de trois réacteurs placés en cascade (fig II.1). Le premier réacteur appelé décanteur primaire reçoit l'eau polluée provenant des milieux urbains ou industriels. C'est là que s'effectue la première phase du traitement qui consiste à éliminer la pollution en suspension et ne laisser passer que de la pollution dissoute. L'eau pénètre ensuite dans un second réacteur appelé bassin d'aération qui constitue le cœur de la station. C'est dans ce dernier que s'opère l'essentiel du traitement par boues activées. Enfin, l'eau subit un dernier traitement dans le troisième réacteur appelé décanteur secondaire. Celui-ci délivre une eau épurée après décantation des boues. Une partie de ces boues est recyclée judicieusement dans le bassin d'aération (boues de retour). L'autre partie (boues en excès) est évacuée pour un traitement extérieur.

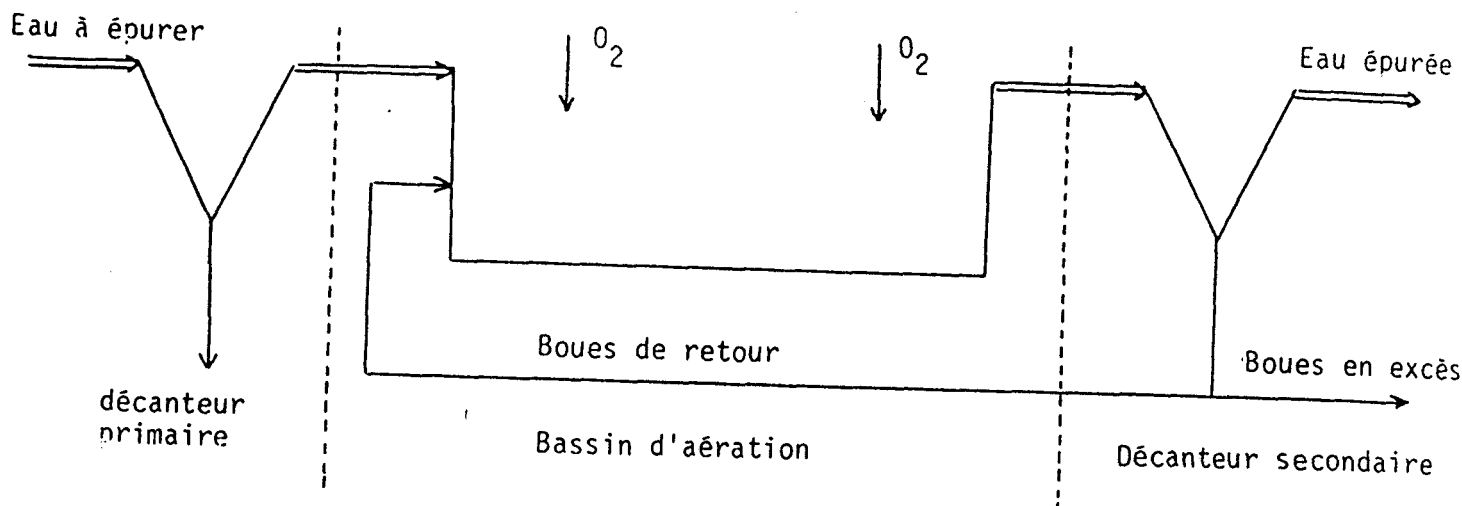


Figure II.1 : Schéma synoptique d'une station d'épuration

V.2 Principe de fonctionnement du système

Après la description physique du système, il convient de décrire son principe de fonctionnement. C'est dans le bassin d'aération que s'effectue le contact entre tous les éléments qui rentrent en jeu (pollution dissoute, micro-organismes, oxygène). Celui-ci mérite une présentation plus approfondie. En effet, cette rencontre donne naissance, si le milieu est favorable, à une prolifération microbienne. La fourniture d'oxygène au milieu est indispensable pour accélérer la multiplication des micro-organismes et faciliter le métabolisme. Cette multiplication se fait progressivement, par division d'une cellule mère en deux cellules filles : c'est le phénomène de croissance. Cette synthèse s'explique par le fait que les micro-organismes ont eu à leurs dispositions des éléments nutritifs en quantité suffisante pour l'édification de nouvelles cellules. L'élément indispensable à cette croissance est principalement le carbone. D'autres éléments interviennent également mais en faible quantité ; on peut citer : l'azote, le sodium, le soufre, ... etc.

Le métabolisme qui est la résultante des transformations précédentes se traduit par une augmentation de la bio-masse. Cette cinétique de métabolisation s'effectue à des vitesses différentes selon :

- la nature du milieu où se pratique l'épuration
- la nature des micro-organismes ;
- la nature du substrat ;
- les réactions bio-chimiques des substrats organiques ;
- la qualité du mélange eau-boues-air.

L'eau usée issue d'un milieu urbain contient principalement des sels nutritifs. Par contre, l'eau provenant d'un milieu industriel est riche en azote et en phosphore. A la limite, on peut dire que la pollution est caractérisée par son contenu énergétique qui peut :

- ☛ mettre à la disposition des micro-organismes, des éléments indispensables à l'édification des nouvelles cellules ;
- ☛ mettre à leur disposition un composé lentement bio-dégradable ou à la limite toxique.

Par conséquent, selon la nature de la pollution, la croissance des micro-organismes obéit à des cinétiques différentes.

Un autre élément joue un rôle très important : c'est le transfert d'oxygène dans l'eau. En effet, la diffusion d' O_2 de l'air dans l'eau permet d'une part de maintenir les micro-organismes en survie (respiration endogène) et d'autre part, de leur donner l'énergie nécessaire pour l'oxydation des matières organiques. C'est pourquoi, le dispositif d'aération doit être judicieusement choisi et placé de façon à remplir efficacement son rôle et éviter les dépôts et la décantation des boues possédant encore une possibilité d'action biologique.

V.3 Les grandeurs d'entrée et de sortie

a) Caractérisation des entrées :

Les entrées à considérer sont les débits d'eau, de boues et leurs concentrations respectives. On ne tient pas compte du débit d'air, on suppose que le bassin est suffisamment oxygéné. Soit donc :

- . D_e, D_b : les débits d'eau et de boues ;
- . S_o, X : les concentrations en substrat en en micro-organismes.

b) Caractérisation de l'état :

Le système étant complexe, il fait apparaître deux états qui interagissent entre eux de façon non simple : l'état hydraulique et l'état biologique.

- L'état hydraulique : La connaissance de cet état reste encore incomplète. Plusieurs recherches ont été faites dans ce sens [LELAN 1972, EL-MALEH 1975, GHOUALMI 1978]. Les auteurs aboutissent à des modèles difficiles à mettre en œuvre sur le plan pratique. Ceci tient à la complexité des phénomènes qui se développent et qui font intervenir à la fois la mécanique des fluides (écoulements triphasiques) et les réactions biochimiques. A ce problème complexe, vient se superposer celui des systèmes d'aération qui diffèrent d'un bassin à l'autre. Pour résoudre en partie ces problèmes, nous avons présenté une méthode de simulation numérique dans le cas d'un mélange PISTON en tenant compte des différents débits ainsi que leurs variations [BARBOUCHA 1982]. C'est sous cette hypothèse que nous étudierons l'aspect modélisation hydraulique du bassin.

- L'état biologique : Cet état fait intervenir l'action biologique du système épurateur. Celle-ci doit tenir compte des inter-actions entre le substrat et les micro-organismes. Une modélisation de cet état doit faire intervenir un modèle du type :

$$\frac{dS}{dt} = f(S, X, \theta)$$

où S : concentration en substrat
 X : concentration en micro-organismes
 θ : un vecteur de paramètres.

c) Caractérisation des sorties

Les sorties à considérer sont : la concentration en pollution en sortie du bassin et son débit correspondant et notées respectivement S_S et D_S . La représentation globale du système est représentée ci-dessous.

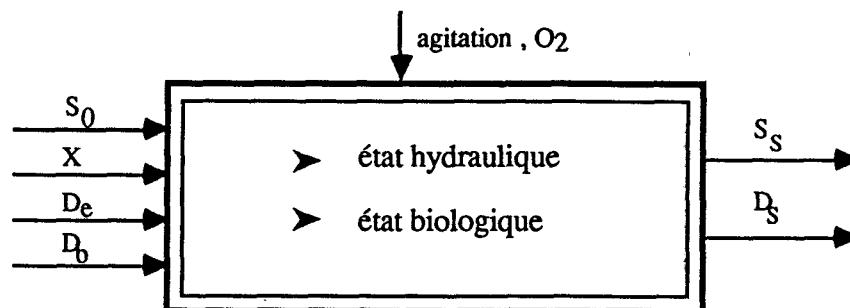


Fig. II.2 : représentation entrée-sortie du système

Parmi toutes ces grandeurs, seuls les débits peuvent être mesurés en temps réel. Ni les différentes concentrations, ni les différents états ne sont mesurables. Nous sommes contraints

d'utiliser un moyen détourné pour estimer les grandeurs non mesurables qui sont indispensables pour une modélisation du système et par conséquent sa régulation. Une telle modélisation doit définir le comportement des micro-organismes devant la pollution introduite. A cet effet, nous avons dû faire appel à un bassin miniature appelé PILOTE.

V.4 Pilote et bilan des grandeurs accessibles

a) Le pilote : n'est rien d'autre qu'un petit bassin d'aération de laboratoire [LAURENT 1984]. Il est constitué d'une cuve d'une capacité de 6 à 10 litres dans laquelle est injectée, en présence des micro-organismes, de l'eau polluée prélevée en amont du bassin. Le schéma synoptique de cette unité est représenté à la figure II.3. Un dispositif d'aération permet d'injecter de l'air en permanence dans la cuve en circulant à contre courant. Un mélange intégral entre les différents éléments est obtenu grâce à une pompe prévue à cet effet. Enfin, une sonde à oxygène mesure en continu la concentration en oxygène dissous $C(t)$ à partir de laquelle on détermine la respiration de micro-organismes $R(t)$.

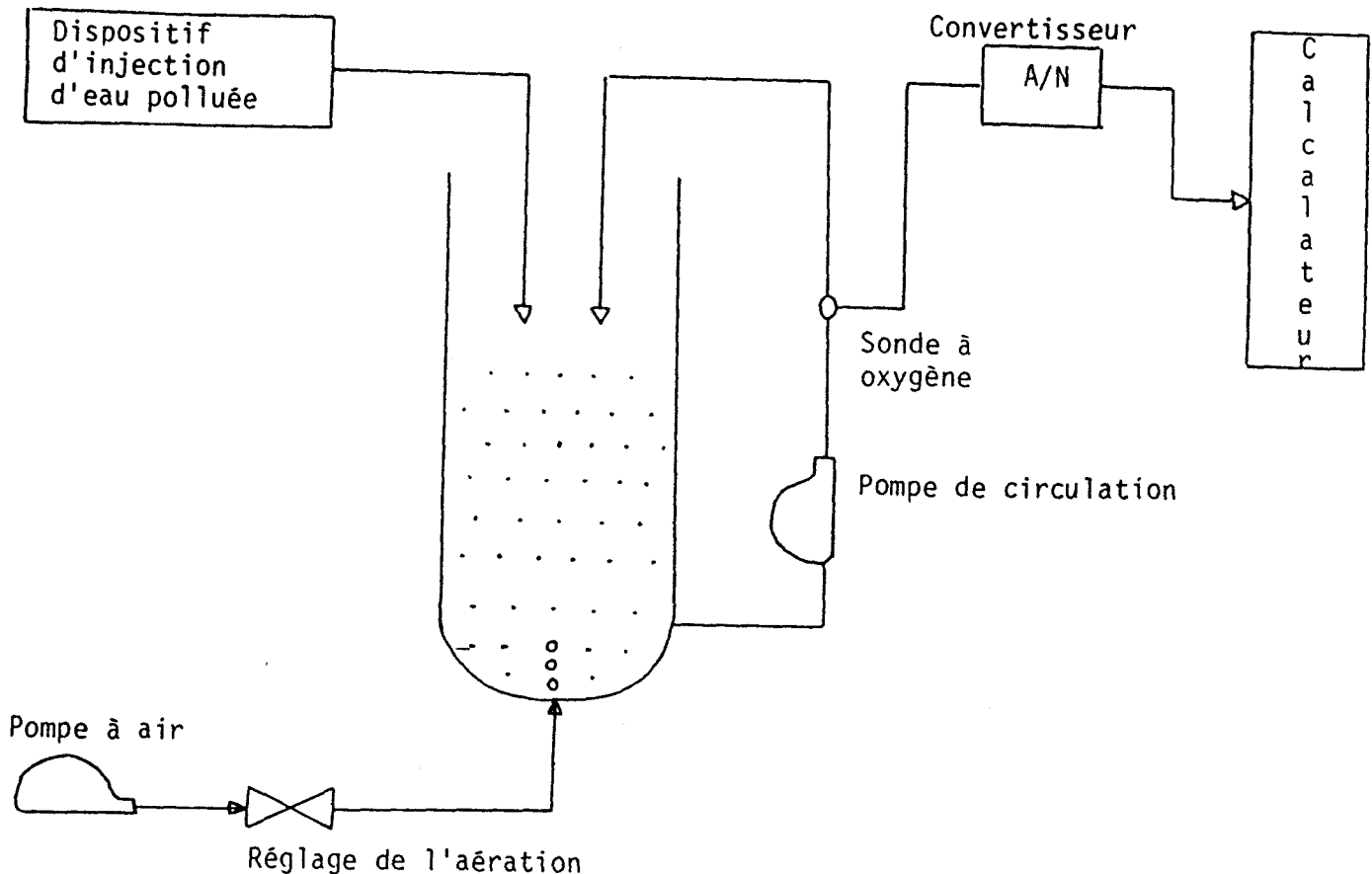


Figure II.3 : Le pilote du CRODA

Le signal issu de la sonde donne la variation de la concentration en oxygène dissous $C(t)$. La détermination de la respiration des micro-organismes $R(t)$ qui est la vitesse de consommation de l'oxygène nécessaire à la dégradation de la pollution introduite, nécessite l'utilisation de la loi classique de variation de $C(t)$ donnée par :

$$dC(t)/dt = K_{LA} (C_{SAT} - C(t)) - R_e - R(t) \text{ avec : } R(t) = a dS(t)/dt$$

(rapport constant entre substrat dégradé et oxygène consommé)

dans laquelle R_e , K_{LA} , C_{SAT} sont des paramètres que nous préciserons dans le tableau ci-dessous et qui peuvent être déterminés facilement dans les conditions suivantes :

. avant l'introduction de la pollution $R(t) = 0$, à l'aide du régime permanent, de la dérivée à l'origine de $C(t)$ et en coupant l'aération en régime permanent on détermine aisément les grandeurs K_{LA} , C_{SAT} et R_e .

La connaissance de ces paramètres ainsi que $C(t)$ et dC/dt conduit à la détermination de $R(t)$ qui caractérise l'affinité substrat micro-organismes en présence de l'oxygène dissous dans le pilote. Celle-ci met en jeu un certain nombre de grandeurs ou paramètres qu'il faudrait adapter au bassin réel. En effet, dans le bassin réel, il faut tenir compte de certaines considérations liées à la géométrie et à la cinétique hydraulique de la cuve d'aération (variation des débits, temps de séjour, gradients de concentration).

En définitive, chaque échantillon de pollution ω_i , $i = 1, \dots, L$ (L taille de l'ensemble des échantillons) est caractérisé par un ensemble de paramètres biologiques et hydrauliques. Nous entendons par échantillon, une quantité de pollution entrant dans le bassin à l'instant t et sortant de celui-ci à l'instant $t + T_S$ où T_S est le temps de séjour de l'eau dans la cuve d'aération. L'ensemble des variables calculées ou mesurées sont résumées dans le tableau ci-dessous. Celles-ci ont été regroupées en trois sous-ensembles variables à expliquer, variables explicatives primaires et variables explicatives secondaires candidates.

b) Bilan des grandeurs accessibles

Le tableau ci-dessous, regroupe l'ensemble des grandeurs connues, mesurées directement ou indirectement. Pour faire une distinction entre ces différentes grandeurs, nous "étiquetons" chacune d'elles avec l'une des lettres suivantes :

- . A : grandeur mesurée à l'aide d'un capteur : mesure directe
- . B : grandeur caractérisant l'évolution temporelle de la concentration en oxygène dissous dans le pilote (donnée par un capteur) : mesure indirecte
- . C : grandeur connue par ailleurs

1 - Les variables à expliquer : $Y = \{X_i, i = 1, \dots, 3\}$	
$X_1 = T_S$: le temps de séjour de l'eau polluée dans le bassin en mn	A
$X_2 = S_S$: concentration résiduelle en pollution à la sortie du bassin en mg/l	A
$X_3 = S_0$: concentration totale en substrat à dégrader en mg/l	A
2 - Les variables explicatives primaires : $X = \{X_i, i = 4, \dots, 15\}$	
$X_4 = KX$: le niveau de saturation des micro-organismes	B
$X_5 = S_0^1$: la quantité du substrat primaire en mg/l	B
$X_6 = b$: vitesse de dégradation des micro-organismes en mg/l.mn	B
$X_7 = \tau$: constante de temps du substrat secondaire en mn	B
$X_8 = S_0^2$: la quantité du substrat secondaire en mg/l	B
$X_9 = \eta$: rapport entre substrat formé et substrat dégradé	B
$X_{10} = Re$: respiration endogène des micro-organismes en mg/l	A
$X_{11} = C_{SAT}$: concentration de saturation en oxygène dissous dans le pilote en mg/l	A
$X_{12} = 1/K_{LA}$: constante de temps de la dissolution de l'oxygène dans l'eau au niveau du pilote en mn	A
$X_{13} = D_e$: débit d'eau polluée en entrée en m ³	A
$X_{14} = D_b$: débit de boues en entrée en m ³	A
$X_{15} = D_S$: débit d'eau polluée en sortie en m ³	A
3 - Les variables explicatives secondaires candidates : $Z = \{X_i, i = 16, \dots, 25, \dots\}$	
$X_{16} = T$: la température du milieu épurateur en 0°C	A
$X_{17} = C$: concentration en oxygène dissous dans le bassin en mg/l	A
$X_{18} = PH$: le PH du milieu épurateur	A
$X_{19} = DCO$: demande chimique en oxygène en mg/l	A
$X_{20} = \mu$: taux de matières en suspension dans l'eau	A
$X_{21} = M$: mode d'aération (intégral, piston ou mixte)	C
$X_{22} = \theta$: taille moyenne des bulles dans le bassin en mn	A
$X_{23} = P_a$: pression atmosphérique en bar	A
$X_{24} = P_A$: perturbations atmosphériques	C
$X_{25} = P_S$: perturbations saisonnières	C
:	:
etc	



Pour plus de détail, on peut se référer à LAURENT [80, 81, 85] et BARBOUCHA [81, 82, 84].

V.5 Modélisation structurale : approche globale et approche locale

Suivant, la variation du débit de pollution et la nature du polluant, nous avons montré dans une étude antérieure que le fonctionnement du bassin d'aération se fait dans l'une des trois zones que l'on peut résumer comme suit [BARBOUCHA 81, LAURENT 85].

- . $S_0 \leq S_{O1}$: faibles charges (effluents peu concentrés)
- . $S_{O1} < S \leq S_{O2}$: charges moyennes
- . $S > S_{O2}$: surcharges (effluents très concentrés, effluents toxiques).

S_0 étant la concentration du substrat d'entrée. Le choix des seuils S_{O1} et S_{O2} est généralement laissé à l'initiative de l'exploitant. Il est bien évident que ces seuils ne seront pas les mêmes pour une station se trouvant dans une zone industrielle et pour une station se trouvant dans une zone urbaine.

V.5.1 Constitution de la population d'apprentissage

Nous avons constitué dans un premier temps une population d'apprentissage de 100 échantillons à partir des données fournies par le C.R.O.D.A. (Centre de Recherche de l'Omnium d'Assainissement). Pour chacun des échantillons, on relève en régime permanent l'ensemble des variables et paramètres le caractérisant. Il s'agit aussi bien de variables hydrauliques et biologiques que des paramètres physiques tels que la température, le PH, ... etc. qui sont résumés dans le tableau ci-dessus. Par ailleurs, chaque variable a été découpée en un certain nombre de modalités. Le choix d'un tel nombre, comme nous l'avons signalé au chapitre I dépend de la nature de chaque variable (variable à expliquer ou variable explicative) et de son domaine de variation [JAMBU 78] [AYGALINC 86]. Dans un tel système, la plupart des variables évoluent très lentement, leur codage ne pose aucun problème particulier. Nous supposons également, pour tester la validité de l'algorithme 2 proposé, que les mesures effectuées sur chaque échantillon ont été répétées une dizaine de fois et qu'un certain nombre d'erreurs auraient pu être commises sur de telles mesures. Une telle hypothèse paraît grossière ; en réalité, il n'en est rien puisque la restitution des données par le pilote nécessite un temps de 15 mn environ et le temps de séjour de l'eau dans le bassin réel est généralement supérieur à 3 heures pour des stations d'épuration de taille moyenne. Par conséquent, l'opérateur a suffisamment le temps de répéter plusieurs fois la même mesure. A l'aide de cette hypothèse, nous avons reconstitué une nouvelle population d'apprentissage de 1 000 échantillons, dont le tableau de contingence est donné en annexe 1. Les lignes d'un tableau sont classées en fonction des valeurs croissantes de S_0 correspondant à la quantité du substrat à dégrader. Ainsi α_1 correspond à un effluent peu concentré et α_{24} est la modalité d'un effluent très concentrée.

V.5.2 Analyse globale et analyse locale

V.5.2.1 Analyse globale

Nous avons utilisé dans un premier temps X comme ensemble explicatif. Ceci nous conduit à générer le tableau de contingence I à partir du tableau initial de données contenant l'ensemble des échantillons Ω .

Le calcul de l'entropie conditionnelle de Y sachant X donne :

$$H(Y/X) = 1,5128 \text{ ou sous forme normalisée } H(Y/X) / H(Y) = 0,38$$

L'entropie résiduelle (ou le % non expliqué de Y) étant élevée, on conclue alors qu'il y a un manque de variables explicatives.

En effet, l'utilisation des tests, conduit à isoler 10 modalités de X (α_{15} à α_{24}) qui ne peuvent expliquer à elles seules les 9 modalités de Y (β_8 et β_{16}). Par conséquent, le recours à des variables explicatives supplémentaires est nécessaire. A cet effet, l'utilisation de l'algorithme 2 conduit à la sélection de l'ensemble explicatif $S^* = \{X_{16}, X_{17}, X_{18}\}$ parmi l'ensemble des variables candidates. Si, on réfère au tableau de contingence II relatif au croisement de Y et X U S^* , l'utilisation des tests permet de conclure que Y est totalement expliqué par X U S^* au bruit près. Or, le calcul de l'entropie résiduelle et le % non expliqué de Y donne :

$$H(Y / X U S^*) = 0,98 \text{ et } [H(Y / X U S^*) / H(Y)] = 0,25$$

Une analyse de modélisation basée uniquement sur ces indices peut conclure à tort l'absence d'une explication totale de Y par X U S^* d'où la nécessité de faire appel aux tests développés dans ce chapitre.

Dans la mesure où le système possède trois zones de fonctionnement, nous allons chercher une partition de Ω en trois classes sous les contraintes suivantes :

$\Omega_1 = \{\omega_i, i \in (1, \dots, 1\ 000) \text{ tel que : } \omega_i \text{ ne possède aucun niveau de saturation}\} : (\text{fig II.5})$

$\Omega_2 = \{\omega_i, i \in (1, \dots, 1\ 000) \text{ tel que : } \omega_i \text{ possède un seul niveau de saturation}\} : (\text{fig. II.6})$

$\Omega_3 = \{\omega_i, i \in (1, \dots, 1\ 000) \text{ tel que : } \omega_i \text{ possède au moins 2 niveaux de saturation}\} : (\text{fig. II.7})$

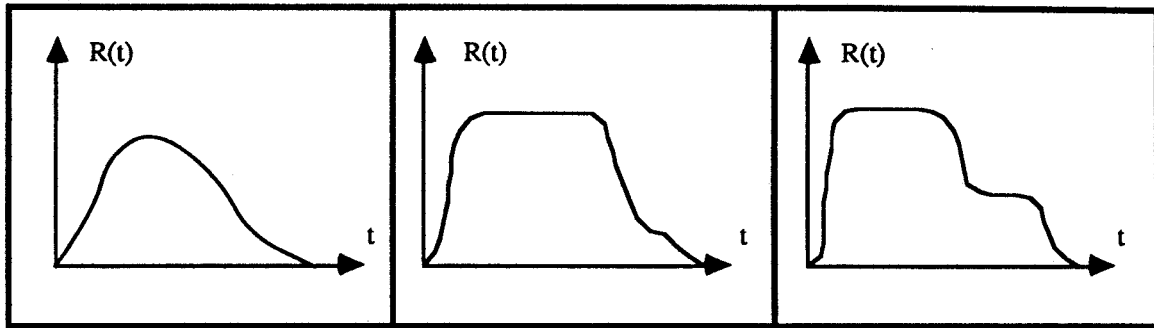


Fig. II.5

Fig. II.6

Fig. II.7

d'où $\Omega_1 = \{\omega_i, i = 1, \dots, 400\} \Rightarrow$ Tableau III

$\Omega_2 = \{\omega_i, i = 401, \dots, 775\} \Rightarrow$ Tableau IV

$\Omega_3 = \{\omega_i, i = 776, \dots, 1\ 000\} \Rightarrow$ Tableau V

On va effectuer alors une analyse locale.

V.5.2.2 Analyse locale

a) Analyse de Ω_1 : (échantillons faiblement concentrés)

Dans ce cas, X explique Y de façon redondante. En effet Y est également expliqué par un sous-ensemble $S \in \mathcal{P}(X)$ de cardinal plus faible. L'utilisation des tests sur le tableau III, montre que Y notée Y^1 est expliqué par S noté X^1 au bruit près alors que, comme dans le cas précédent, le calcul de l'entropie résiduelle conclut à une insuffisance d'explication de Y^1 par X^1 puisque :

$$H(Y^1/X^1) = 1,27 \quad \text{et} \quad [H(Y^1/X^1) / H(Y^1)] = 0.41$$

avec $X^1 = X \setminus \{X_4, X_7, X_8, X_9\} = S$

X^1 modélise Y^1 au bruit près, ceci s'explique par le fait que pour de tels effluents, il y a absence de saturation des micro-organismes d'une part et la dégradation du substrat primaire ne génère pas de substrat secondaire d'autre part. Cela apparaît clairement sur la courbe $R(t)$ (fig. II.5) qui ne présente pas les deux caractéristiques précédentes. Pour de tels échantillons, il n'est pas nécessaire de faire appel à un modèle de dimension importante pour décrire le système dans cette zone de fonctionnement.

b) Analyse de Ω_2 : (cas des charges moyennes)

L'analyse d'un tel échantillon, nous a permis de conclure que Y est expliqué par X et ne peut être expliqué par un sous-ensemble de cardinal plus faible que X. Dans ce cas, on peut re-

marquer que la présence de saturation des micro-organismes ainsi que l'apparition du substrat secondaire (fig. II.6) justifie la présence des variables X_4 , X_7 , X_8 et X_9 parmi les variables constituant l'ensemble explicatif X .

Le calcul de l'entropie résiduelle à partir du tableau IV donne :

$$H(Y^2/X^2) = 0,91 \quad \text{et} \quad [H(Y^2/X^2) / H(Y^2)] = 0,34$$

même remarque que précédemment sur l'utilisation de ces indices pour juger l'explication de Y^2 par X^2 .

c) Analyse de Ω_3 (cas des fortes charges ou charges toxiques)

La modélisation d'un tel sous-échantillon, nécessite l'utilisation d'un sous-ensemble explicatif supplémentaire $S^* = \{X_{16}, X_{17}, X_{18}\}$. En effet, c'est ce sous-échantillon qui est l'origine de l'absence de l'explication totale de Y sur l'ensemble de ses modalités.

L'analyse du tableau V montre que Y noté Y^3 est expliqué par $X \cup S^*$ noté X^3 au bruit près même si :

$$H(Y^3/X^3) = 0,52 \quad \text{et} \quad [H(Y^3/X^3) / H(Y^3)] = 0,23$$

La modélisation du système dans cette zone de fonctionnement est plus complexe que dans les deux cas précédents pour les raisons suivantes :

- ☞ il y a présence d'un substrat primaire dont la dégradation donne naissance à un substrat secondaire ; chacun d'eux possède un niveau de saturation ;
- ☞ la complexité de tels effluents conduit à une variation de paramètres tels que : le PH, la température, la concentration en oxygène dissous dans le bassin (variables X_{16} , X_{17} et X_{18}) qui sont en général constants pour des effluents moyennement ou faiblement concentrés [LAURENT 85].

En définitive, l'approche globale conduit à la recherche d'un modèle qui est certes complet, mais décrit de façon redondante certains types d'échantillons qui ne nécessitent qu'un modèle de taille plus réduite. Par contre, l'approche locale, permet de proposer pour chaque zone de fonctionnement du système, un modèle partiel permettant de décrire de façon satisfaisante son comportement dans une telle zone. Ainsi, dans le système que nous avons analysé ; pour la modélisation des échantillons de Ω_1 , on peut faire appel au modèle classique de croissance microbienne de MONOD [1942], alors que les échantillons Ω_2 et Ω_3 nécessitent un modèle plus général. Le modèle de LAURENT-POVY [1978] semble convenir pour décrire de tels échantillons.

VI - CONCLUSION

L'analyse des relations entre variables, à partir des données relevées sur le système étudié, constitue souvent le premier pas vers une modélisation de celui-ci. Pour éviter d'augmenter inutilement la dimension du modèle recherché, il convient d'étudier, pour chaque variable explicable, ses sous-ensembles explicatifs optimaux. Nous avons montré, certaines propriétés des sous-ensembles explicatifs qui permettent d'accélérer notablement le processus de leur détermination dans le cas d'un système déterministe [ALGORITHME 1].

Sous un angle plus général, nous avons montré l'insuffisance des approches classiques pour la faisabilité du problème de modélisation lors de la présence du bruit sur les variables du système. Pour résoudre un tel problème, nous avons fait appel aux tests statistiques et avons développé un indice basé sur la variation entropique. Cette façon de procéder, nous a permis de déceler les modalités non expliquées des variables que l'on se propose de modéliser.

L'utilisation des algorithmes de type hiérarchique, pour la recherche des variables explicatives supplémentaires parmi les variables candidates conduit à l'explication du système [ALGORITHME 2]. A partir de la donnée des relations validées, il est alors possible de construire le graphe d'interconnexions et de mettre en évidence, par sa décomposition, sa structure hiérarchisée pour obtenir celle du système [RICHETIN 75, MILGRAM 75].

Enfin, nous avons développé une application de l'approche proposée sur un système complexe : les stations d'épuration des eaux usées et avons montré l'utilité de la modélisation structurale. Nous avons également mis en évidence l'intérêt d'une approche locale pour décrire les fonctionnements multiples éventuels d'un système.

CHAPITRE III

MODELISATION DES SYSTEMES
COMPLEXES
APPROCHE SYSTEME EXPERT

I - INTRODUCTION

Les systèmes complexes sont souvent constitués d'un nombre important d'éléments qui interagissent entre eux de façon non simple [SIMON 1962]. Leur prise en compte, pose de sérieux problèmes de modélisation et d'analyse de relations fonctionnelles difficilement maîtrisables par les approches classiques. Nous avons montré que l'approche modélisation structurale utilisant les concepts de la théorie de l'information peut apporter une réponse à cette nécessité. Les systèmes suffisamment complexes, sur lesquels on ne dispose d'aucune information a priori et qui ne peuvent être appréhendés au moyen de relations fonctionnelles nécessitent une procédure d'analyse permettant d'extraire les connaissances qu'ils recèlent. Par ailleurs, une méthode d'analyse est jugée par sa flexibilité, son adaptation à tout type de systèmes et par la manière dont elle utilise les connaissances. Le choix d'une méthode, l'interprétation des résultats nécessite en effet la possibilité de manipuler des connaissances, un savoir souvent de type déclaratif qui n'est pas toujours possible par des programmes classiques procéduraux [DEMONCHAUX 1985]. C'est pourquoi, un grand nombre de recherches laissant de côté l'approche algorithmique traditionnelle s'orientent vers des techniques d'intelligence artificielle et particulièrement les systèmes experts. La caractéristique principale de cette approche est la séparation introduite entre base de connaissances et mécanismes d'exploitation de ces connaissances [PINSON 1981] [LAURIERE 1982] [LAURENT 1984]. En effet, la mise en œuvre de tels systèmes nécessite leur décomposition en trois composantes fondamentales :

- la base de connaissance qui intègre le savoir acquis sur le long terme par l'expert ;
- la base des faits qui contient l'ensemble des données descriptives du domaine exploré ;
- le moteur d'inférence qui est le mécanisme d'exploitation ou le programme interpréteur.

L'avantage de ces systèmes, par rapport à l'approche algorithmique traditionnelle, réside dans le fait qu'ils utilisent des heuristiques acquises auprès des experts humains, ce qui leur permet de "tracer" le raisonnement utilisé en des termes compréhensibles par les utilisateurs. Ils présentent également d'autres avantages, en plus de la séparation introduite entre la connaissance et les mécanismes chargés de l'exploiter qui constituent un facteur primordial de confiance pour les utilisateurs, tels que : l'évolutivité, l'explication du raisonnement, ... et permettent ainsi d'automatiser des tâches pour lesquelles il est difficile d'établir des algorithmes de résolution [TALLIBERT 1985]. Comme toutes les disciplines novatrices, les systèmes experts utilisent des méthodes dont le formalisme et les concepts ne sont pas encore fixés ni bien éprouvés. Y. KODRATOFF [1985] dit à ce propos "Malgré leur popularité, les systèmes experts ne sont pas clairement définis. Ils correspondent plus à un mode d'approche des problèmes qu'à une caractéristique formelle, leur définition varie selon les écoles et selon les individus". Comme le dit également C. GRANGER [1985] "Il n'existe pas à l'heure actuelle de méthodes théoriques générales capables d'indiquer quelle méthode de représentation choisir selon les fonctionnalités

liées au domaine d'application et au type de problèmes à résoudre". Toute la difficulté réside dans l'élaboration de la base des connaissances du problème à analyser. L'exploitation d'une telle base de connaissance a fait l'objet d'un grand nombre de travaux visant à définir des moteurs d'inférence efficaces [CORDIER 1985] [LEBEUX 1986]. Son élaboration, par contre fait appel à un ou plusieurs experts qui traduisent sous forme de règles une expérience acquise sur le long terme. Or, peu d'experts humains sont capables d'exprimer de façon exhaustive toutes les connaissances liées à leurs domaines, de ce fait, les erreurs et omissions peuvent être fréquentes.

En effet, la réalisation d'un système expert, se heurte encore à de nombreux problèmes liés essentiellement à la "transférabilité" du savoir des experts pour laquelle aucune méthodologie n'existe encore. D'autant plus que, comme le signale V. RIALLE [1985], "Il existe quelques domaines pas trop vastes, incertains et incomplets et pour lesquels n'existe pas d'expert à proprement parler capable de construire une base de connaissance utilisable par un moteur d'exploitation. De tels domaines ne nécessitent pas moins une aide au raisonnement et à la déduction". Par ailleurs, un certain nombre de systèmes sont non stationnaires, dans le sens où les règles régissant leur évolution varient en temps "réel" (horizon d'un jour à un an par exemple), ou bien présentent des modes de fonctionnement multiples, régis par différents ensembles de règles. Pour de tels systèmes, toute analyse est basée sur l'acquisition et l'exploitation, sur un horizon d'observation approprié des données qu'ils génèrent. En effet, ces données traduisent le comportement du système et il convient d'en extraire la connaissance qu'elles recèlent. A cet effet, nous proposons deux approches complémentaires pour l'exploitation des tableaux de contingence regroupant les informations disponibles sur le domaine à analyser : la première est basée sur la définition d'un opérateur de contraste appliqué à ces tableaux, la seconde fait appel à un coefficient d'extraction de l'information utilisant les concepts de l'analyse structurale par le biais de la théorie de l'information. Les deux approches proposées seront comparées et utilisées conjointement avec une méthode géométrique d'analyse des données : l'analyse factorielle des correspondances, pour l'extraction et l'expression des connaissances contenues dans un tableau de contingence.

II - NOTATIONS COMPLEMENTAIRES

Le système que l'on se propose d'analyser fait intervenir les sous-ensembles Ω , X , Y , M_X et M_Y définis aux chapitres précédents.

L'analyse du système nécessite la construction du tableau de contingence $[P_{ij}]$ (Fig. III.1) à partir du tableau initial de données. La case (i,j) du tableau contient P_{ij} la probabilité d'occurrence conjointe des modalités α_i et β_j .

Y X	β_1 -----	β_j	----- β_m	
α_1 ⋮ ⋮				
α_i		P_{ij}		$P_{i.}$
⋮ ⋮ α_n				
		$P_{.j}$		

Fig. III.1 : tableau de contingence $[P_{ij}]$

On note par ailleurs :

$$P_{i.} = \sum_j P_{ij} \quad ; \quad P_{.j} = \sum_i P_{ij}$$

$$\sum_i \sum_j P_{ij} = \sum_i P_{i.} = \sum_j P_{.j} = 1$$

$$P_{j/i} = \frac{P_{ij}}{P_{i.}} \quad ; \quad P_{i/j} = \frac{P_{ij}}{P_{.j}}$$

$$B(\alpha_i) = \{\beta_j, j \in (1, \dots, m) \text{ tel que : } P(Y = \beta_j / X = \alpha_i) = P_{j/i} > 0\}$$

$$A(\beta_j) = \{\alpha_i, i \in (1, \dots, n) \text{ tel que : } P(X = \alpha_i / Y = \beta_j) = P_{i/j} > 0\}$$

III - QUELLE CONNAISSANCE PEUT-ON EXTRAIRE DE $[P_{ij}]$

Pour une exploitation judicieuse du tableau $[P_{ij}]$, les deux seules hypothèses requises concernent la taille L de l'échantillon et l'horizon d'observation [Cf § 4.2.6.1. CHAP I].

Dans les chapitres I et II, nous avons montré l'intérêt de l'approche locale pour l'étude d'un système. Dans ce sens, nous proposons dans un premier temps l'exploitation du contenu informationnel de chaque ligne [resp de chaque colonne] du tableau $[P_{ij}]$ que nous présenterons sous forme de règles.

III.1 Définition

Une règle est définie comme l'information extraite sous forme symbolique à partir d'une distribution de probabilité issue du tableau de contingence $[P_{ij}]$.

D'une manière opérationnelle, une règle extraite d'une ligne peut s'écrire comme suit :

$$X = \alpha_i \text{ ALORS } Y = \beta_1 \text{ OU } \dots \text{ OU } \beta_j \text{ OU } \dots \text{ OU } \beta_m$$

Il en est de même pour une règle extraite d'une colonne :

$$Y = \beta_j \text{ ALORS } X = \alpha_1 \text{ OU } \alpha_2 \text{ OU } \dots \text{ OU } \alpha_i \text{ OU } \dots \text{ OU } \alpha_n$$

III.2 Problèmes posés lors de la formulation des connaissances

A l'aide des informations contenues dans $[P_{ij}]$, il est possible de traduire des connaissances sous forme de règles. Mais la difficulté réside dans leur sélection et leur évaluation. Ainsi l'implication $X = \alpha_i \Rightarrow Y = \beta_j$ doit être assortie d'un degré de vérité permettant de mesurer sa concordance avec la réalité observée. L'utilisation d'un coefficient de vérité V (ou de non vérité $N.V.$) basé uniquement sur les concepts probabilistes n'est pas suffisant pour la pondération d'une telle règle. A titre d'exemple, prenons pour $X = \alpha_i$ deux distributions de probabilité P_1 ($Y/X = \alpha_i$) et P_2 ($Y/X = \alpha_i$) (Fig. III.2). Dans les deux cas, on peut affirmer que :

$$. X = \alpha_i \text{ ALORS } Y = \beta_1 \text{ avec } V = 0.7 \text{ et } NV = 0.3$$

Les distributions P_1 et P_2 donnent naissance à la même règle. Pourtant toute la connaissance utilisable à partir de la distribution P_1 n'a pas encore été exploitée alors que manifestement la distribution P_2 ne peut pas fournir d'information supplémentaire.

Cet exemple montre l'insuffisance des degrés de pondération d'une règle conçus seulement à partir des modèles probabilistes. Quelle connaissance peut-on extraire effectivement de P_1 et P_2 ? Dans ce cas précis, on ne s'écarterait pas trop de la réalité en formulant les règles suivantes.

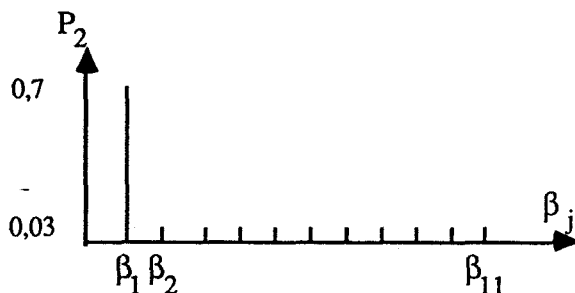
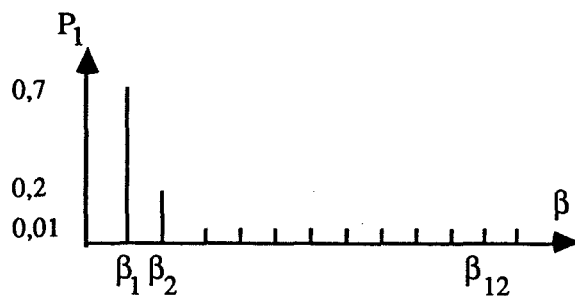


Fig. III.2 : Les distributions de P_1 et P_2

- . cas (a) : $X = \alpha_i$ ALORS $Y = \beta_1$ OU β_2
- . cas (b) : $X = \alpha_i$ ALORS $Y = \beta_1$

Dans les deux cas, on néglige les résidus qui ne peuvent fournir une connaissance opérationnelle.

Voyons à présent quels sont les types de règles susceptibles d'être extraites du tableau de contingence ?

III.3 Nature des règles

Dans l'approche système expert, la représentation de la connaissance sous forme de règles est le formalisme le plus utilisé à cause de sa souplesse et de sa puissance. D'autres approches sont également proposées, elles utilisent les réseaux sémantiques, les frames, ... etc. On se limitera dans ce travail à l'approche la plus adoptée, celle des règles de production. Deux types de règles peuvent être distinguées.

III.3.1 Les règles de type 1

Elles utilisent les informations contenues dans une ligne de $[P_{ij}]$ et s'écrivent comme suit :

$$. X = \alpha_i \text{ ALORS } Y \in B_{\sigma}(\alpha_i) \subseteq B(\alpha_i) \quad (\text{III.1})$$

. dans le cas où $B_{\sigma}(\alpha_i)$ se réduit à un seul élément β_j par exemple, la règle précédente s'écrit :

$$X = \alpha_i \text{ ALORS } Y = \beta_j$$

. dans le cas défavorable où $B_{\sigma}(\alpha_i)$ contient l'ensemble des modalités $\beta_j, j = 1, \dots, m$, c'est-à-dire $B_{\sigma}(\alpha_i) = B(\alpha_i)$, la règle s'exprime alors :

$$X = \alpha_i \text{ ALORS } Y = \beta_1 \text{ OU } \beta_2 \text{ OU } \dots \text{ OU } \beta_m \quad -$$

Le quantificateur logique "ET", que nous noterons " \wedge ", est implicite en hypothèse est en conclusion, ainsi :

$$. X = \alpha_i \text{ signifie : } (X_1 = \alpha_i^1) \wedge (X_2 = \alpha_i^2) \wedge \dots \wedge (X_p = \alpha_i^p) \quad (\text{III.2})$$

$$. Y = \beta_j \text{ signifie : } (Y_1 = \beta_j^1) \wedge (Y_2 = \beta_j^2) \wedge \dots \wedge (Y_q = \beta_j^q)$$

Par ailleurs, l'exploitation de l'information contenue dans deux ou plusieurs lignes du tableau $[P_{ij}]$ (agrégation des lignes) conduit à des règles composées qui peuvent s'exprimer comme suit :

$$X = \alpha_i \text{ OU } \alpha_k \text{ ALORS } Y \in B_{\sigma_i}(\alpha_i) \cup B_{\sigma_k}(\alpha_k) \quad (\text{III.3})$$

Nous remarquons que le quantificateur logique "OU" que nous noterons "V" est explicite en hypothèses (regroupement des lignes) et en conclusion.

III.3.2 Les règles de type 2

Elles font apparaître des coefficients de pondération probabilistes pour chacune des modalités prises par Y en tenant compte de l'exploitation ou la non exploitation de l'information résiduelle. Ainsi, si Y prend une seule modalité β_j , la règle s'écrit :

. $X = \alpha_i \text{ ALORS } Y = (\beta_j, r_{j/i} = 1)$ où $r_{j/i}$ est le coefficient de pondération de la règle. En l'absence des résidus $r_{j/i}$ n'est autre que $P_{j/i}$, dans le cas contraire $r_{j/i} > P_{j/i}$.

D'une manière générale de telles règles s'écrivent comme suit :

$$. X = \alpha_i \text{ ALORS } Y = (\beta_1, r_{1/i}) \vee (\beta_2, r_{2/i}) \vee \dots \vee (\beta_\sigma, r_{\sigma/i}) \quad (\text{III.4})$$

avec $1 \leq \sigma \leq m$; règle qui signifie qu'à chaque modalité $\beta_j, j \in \{1, \dots, \sigma\}$ est associé un coefficient de pondération $r_{j/i}$ contrairement aux règles de type 1 où ces coefficients ne sont pas spécifiés.

Chaque ligne [resp chaque colonne] du tableau $[P_{ij}]$ donne naissance à une règle. Au total une liste de $(n + m)$ règles peuvent alors être extraites, elles constituent le modèle redondant du système. Le problème de regroupement et de simplification de telles règles, en vue de la génération du modèle simplifié du système sera étudié au chapitre IV.

Dans cette partie, notre intérêt est porté sur la façon dont les informations disponibles dans le tableau $[P_{ij}]$ peuvent être exploitées et avec quelle précision des règles peuvent être déduites : à cet effet, nous proposons deux approches complémentaires.

IV - EXTRACTION DE CONNAISSANCES : APPROCHE UTILISANT UN OPERATEUR DE CONTRASTE

C'est une approche locale qui consiste à identifier les cases importantes de chaque ligne [resp colonne] du tableau $[P_{ij}]$ et à utiliser au mieux son contenu informationnel. Ainsi, si on découvre par exemple qu'une case renferme 95 % de l'information contenue dans une ligne, il se-

rait inutile de conserver les $(m - 1)$ autres cases qui n'apporteront qu'une information négligeable. D'où l'intérêt de déceler les cases qui sont susceptibles d'apporter une information exploitable. A cet effet, nous allons définir dans un premier temps un opérateur de contraste qui nous permettra d'extraire les informations pertinentes contenues dans le tableau $[P_{ij}]$.

IV.1 Opérateur de contraste

. Considérons une ligne $i, i \in \{1, \dots, n\}$ du tableau $[P_{ij}]$ et soit $P (Y/X = \alpha_i)$ sa distribution conditionnelle de probabilité dont les éléments sont supposés ordonnés ; c'est-à-dire :

$$P (Y/X = \alpha_i) = \{P_{j/i}, j = 1, \dots, m \text{ tel que : } P_{1/i} \geq P_{2/i} \geq \dots \geq P_{m/i}\} \quad (\text{III.5})$$

. On cherche à élaborer, à partir de la distribution $P (Y/X = \alpha_i) = P$, pour une troncature σ fixée, $1 \leq \sigma \leq m$, une nouvelle distribution $R (Y/X = \alpha_i) = R$ aussi proche que possible de P (Fig. III.3).

. La distribution optimale R^* doit être choisie telle que $D (P, R^*)$ est minimal où $D (P, R^*)$ est un indice de distance entre les deux lois de probabilité.

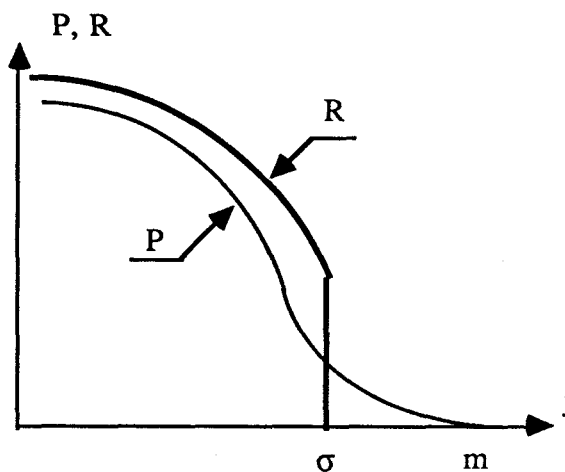


Fig. III.3 : Tracé de R à partir de P et pour une troncature σ donnée.

De nombreux indices existent dans la littérature et permettent la comparaison des lois de probabilité : distance du χ^2 (noté D_1), l'information discriminante de KULLBACK (D_2), le χ^2 de NEYMAN'S (D_3) et l'information de SHANNON (D_4) [BENZECRI 1980] [KULLBACK 1962].

Nous montrons (Cf. Annexe III.1) que la distribution R^* suivante est optimale pour les quatre indices ci-dessus :

$$R^* (Y/X = \alpha_i) = \left\{ r_{j/i}, j = 1, \dots, \sigma \text{ tel que : } r_{j/i} = P_{j/i} + P_{j/i} \frac{\sum_{\sigma+1}^m P_{j/i}}{\sum_1^{\sigma} P_{j/i}} = \frac{P_{j/i}}{\sum_1^{\sigma} P_{j/i}} \right\}$$

avec $r_{j/i} = 0$ pour $j = \sigma + 1, \dots, m$ (III.6)

L'information résiduelle perdue est répartie de façon proportionnelle sur les raies conservées.

Remarque IV.1.1.

L'utilisation d'un indice de distance basé uniquement sur le calcul de la différence entropique entre les deux distributions : $\Delta H = H(P) - H(R^*) = D_5$, conduit à R^* optimale qui s'écrit :

$$R^* = \{r_{j/i}, j = 1, \dots, \sigma \text{ tel que } r_{j/i} = 1/\sigma \text{ avec } r_{j/i} = 0, \forall j = \sigma+1, \dots, m\} \quad (\text{III.7})$$

En effet $H(P)$ est constant, ΔH est minimum lorsque $H(R)$ est maximum. Celui-ci est obtenu dans le cas où les modalités $\beta_j, j = 1, \dots, \sigma$ sont équiprobables.

Cette situation correspond particulièrement au cas des règles de type 1 où la probabilité de chacune des modalités n'est pas spécifiée. Autrement dit, elles sont également représentées ou bien uniformément distribuées sur l'intervalle $\{1, \dots, \sigma\}$.

Remarque IV.1.2.

L'utilisation d'un indice du type quadratique qui s'écrit :

$$D(P, R^*) = D_6 = \sum_1^m |r_{j/i} - P_{j/i}|^\gamma \quad \text{où } \gamma > 0$$

conduit à la définition de R^* optimale comme suit (démonstration annexe III.1) :

$$R^* (Y/X = \alpha_1) = \left\{ r_{j/i}, j = 1, \dots, \sigma \text{ tel que } r_{j/i} = \left(P_{j/i} + \frac{\sum_{\sigma+1}^m P_{j/i}}{\sigma} \right) \right\}$$

avec $r_{j/i} = 0$ pour $j = \sigma+1, \dots, m$ (III.8)

Dans ce cas les raies conservées, se trouvent augmentées de la même quantité :

$$\frac{\sum_{\sigma+1}^m P_{j/i}}{\sigma}$$

Mis à part ces deux cas particuliers, le choix de R^* donné par (III.6) qui minimise les indices les plus connus nous conduit à définir un opérateur que nous appellerons : opérateur de contraste comme suit :

a) Définition : L'opérateur de contraste C_σ , de niveau σ relatif à une ligne du tableau $[P_{ij}]$ est :

$$C_\sigma = \begin{cases} \frac{1}{\sigma} & \text{pour } j \in \{ 1, \dots, \sigma \} \\ \sum_1^\sigma P_{j/i} & \\ 0 & \text{pour } j \in \{ \sigma+1, \dots, m \} \end{cases}$$

. Contraster une ligne i du tableau consiste à multiplier chacun des éléments $P_{j/i}$, $j = 1, \dots, m$ par son coefficient de contraste.

. Si on note D l'ensemble des distributions conditionnelles de probabilité, la définition formelle de C_σ est donnée par :

$$C_\sigma : D \longrightarrow D$$

$$P(Y/X = \alpha) \longrightarrow R^*(Y/X = \alpha) = C_\sigma [P(Y/X = \alpha)]$$

. Ceci montre qu'une fois les éléments de la ligne contrastés on obtient une nouvelle distribution $R^*(Y/X = \alpha)$, que nous noterons désormais $R(Y/X = \alpha)$, définie par : $R(Y/X = \alpha) = C_\sigma [P(Y/X = \alpha)] = C_\sigma \cdot P(Y/X = \alpha)$

. Les éléments de cette nouvelle distribution sont également ordonnés (ordre décroissant), l'ordre des σ premiers éléments étant inchangé :

$$R(Y/X = \alpha) = \{r_{j/i}, j = 1, \dots, \sigma \text{ tel que : } r_{1/i} \geq \dots \geq r_{\sigma/i}\} \quad (\text{III.11})$$

. Il s'agit effectivement d'une distribution de probabilité, en effet on a :

$$\sum_{j=1}^m r_{j/i} = \sum_{j=1}^{\sigma} r_{j/i} = 1$$

. Par ailleurs, l'opérateur de contraste sur les lignes conserve leur distribution marginale :

$$r_{i.} = P_{i.}$$

. Pour un niveau de contraste σ donné, un élément j de la ligne i s'écrit :

$$R (Y=\beta_j / X=\alpha_i) = \begin{cases} r_{j/i} = \frac{P_{j/i}}{\sum_1^{\sigma} P_{j/i}} = \frac{P_{ij}}{\sum_1^{\sigma} P_{ij}} ; \quad \forall j \in \{ 1, \dots, \sigma \} \\ r_{j/i} = 0 ; \quad \forall j \in \{ \sigma+1, \dots, m \} \end{cases} \quad (\text{III.12})$$

On appellera résidus les quantités $P_{j/i}$, $j \in \{ \sigma + 1, \dots, m \}$ qui correspondent aux modalités éliminées par l'opérateur de contraste.

b) Limites de C_{σ}

. si $\sigma = m$ alors $c_{\sigma} = 1$; les deux distributions sont identiques :

$R (Y/X = \alpha) = P (Y/X = \alpha)$: dans ce cas aucun contraste n'est effectué.

. si $\sigma = 1$ alors $C_{\sigma} = 1/P_{1/i} \Rightarrow R (Y/X = \alpha) =$

1 pour $j = 1$
0 pour $j = 2, \dots, m$

Dans ce cas, un contraste maximum est effectué.

Par conséquent, dans le cas général :

$$1 \leq C_{\sigma} \leq \frac{1}{P_{1/i}} \quad (\text{III.13})$$

c) Exemple :

Soit $P (Y/X = \alpha) = \{ 0.4, 0.35, 0.15, 0.01, \dots, 0.01 \}$ avec $\sigma = 3$ et $P_i = 0.1$. Le calcul de C_{σ} donne :

$$C_{\sigma} = \begin{cases} \frac{0,1}{0,09} = \frac{10}{9} & \text{pour } j = 1, 2, 3 \\ 0 & \text{pour } j = \{ 4, \dots, m \} \end{cases}$$

La nouvelle distribution s'écrit : $R (Y/X = \alpha) = R = \{ 4/9, 3.5/9, 1.5/9, 0, \dots, 0 \}$.

Nous avons tracé à la figure III.4 les deux distributions P et R ainsi que C_{σ} .

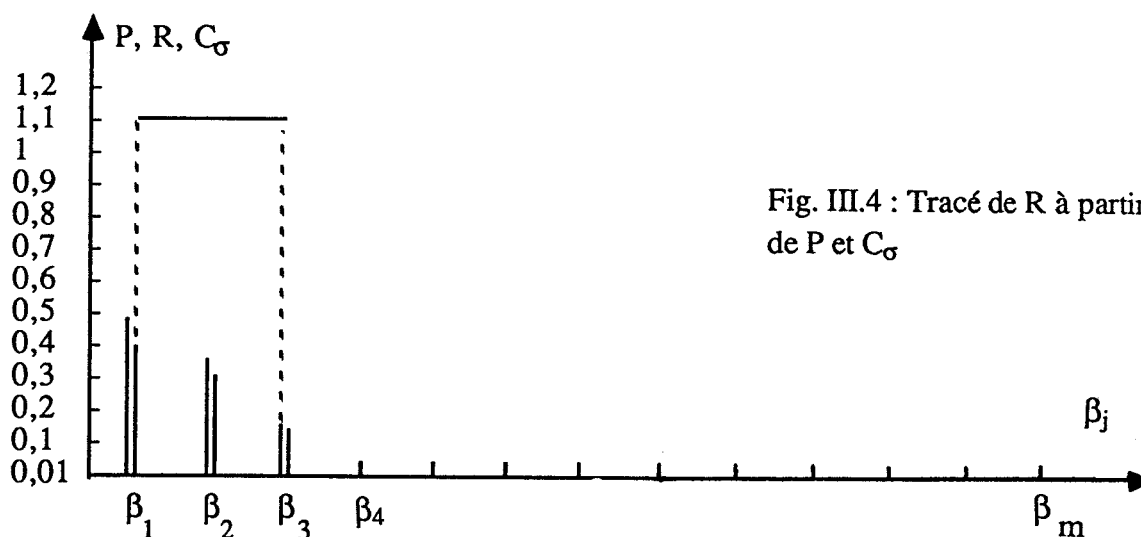


Fig. III.4 : Tracé de R à partir de P et C_σ

Remarque IV.1.3.

Il est possible de définir un opérateur de contraste quadratique. En effet, dans le cas précédent (contraste linéaire), les σ éléments de la distribution sont multipliés par un coefficient constant égal à :

$$\frac{1}{\sum_1^\sigma P_{j/i}}$$

La multiplication de chaque élément par un coefficient qui lui est proportionnel conduit à un opérateur de contraste quadratique (cf. annexe III.2). Celui-ci renforce les valeurs les plus élevées de la distribution et diminue celles qui sont faibles. On se limitera dans ce travail au cas linéaire.

IV.2 Analyse du niveau de contraste

Le problème qui se pose à ce stade est celui de la détermination du niveau de contraste C_σ à effectuer sur chaque ligne du tableau $[P_{ij}]$ de sorte que l'information perdue lors de cette opération reste minimale.

Pour résoudre un tel problème, on peut faire appel dans un premier temps, à un indice de distance faisant apparaître le coefficient de contraste ; soit :

$$d_1(\sigma, P) = C_\sigma - 1 = \frac{1}{\sum_1^\sigma P_{j/i}} - 1 = \frac{1 - \sum_1^\sigma P_{j/i}}{\sum_1^\sigma P_{j/i}} = \frac{\sum_1^m P_{j/i}}{\sum_1^\sigma P_{j/i}} \quad (\text{III.14})$$

$d_1(\sigma, P)$ fait apparaître le rapport entre la probabilité des résidus éliminés et la probabilité des éléments conservés après contraste.

La perte d'information occasionnée par cette opération peut être également déterminée à l'aide des concepts entropiques.

Notons $H_P (Y/X = \alpha)$ [respectivement $H_R (Y/X = \alpha)$] l'entropie de Y sachant que X prend la modalité α calculée avec la loi de distribution conditionnelle P [resp R]. $H_P (Y/X = \alpha) = H_P$ s'interprète de la même façon, ainsi la perte d'information ΔH occasionnée par le remplacement de P par R s'écrit :

$$\Delta H = H_P - H_R = - \sum_1^m P_{j/i} \text{Log } P_{j/i} + \sum_1^\sigma r_{j/i} \text{Log } r_{j/i} \quad (\text{III.15})$$

$$= - \sum_1^m P_{j/i} \text{Log } P_{j/i} + \frac{1}{\sum_1^\sigma P_{ij}} \cdot \sum_1^\sigma P_{ij} \text{Log } P_{ij} - \text{Log } \sum_1^\sigma P_{j/i}$$

$$\text{ou, sous forme normalisée : } \frac{\Delta H}{H_P} = \frac{H_P - H_R}{H_P} = d_2(\sigma, P) \quad (\text{III.16})$$

Cette quantité représente le pourcentage de l'information perdue à l'issue de l'opération de contraste et le niveau de contraste à retenir peut être déterminé en se fixant un seuil de perte d'information limite.

Un tel problème peut être également formulé comme un problème de test statistique. En effet, P est une distribution réelle observée sur le sous-échantillon L_α de L caractérisé par $X = \alpha$, R est une distribution théorique censée décrire d'une manière plus exploitable ce même échantillon. Le problème posé est alors, de savoir jusqu'à quel point ces deux distributions représentent la même réalité. Autrement dit, il s'agit de tester pour un σ donné, les deux hypothèses suivantes :

- H_0 : les distributions P et R sont approximativement identiques ;
- H_1 : les distributions P et R sont significativement différentes.

Ce type de problème, largement étudié dans la littérature [BENZECRI 1980] [SAPORTA 1978] [KULLBACK 1962] reçoit une simple et classique solution par la mise en œuvre des tests statistiques suivants :

- le test du χ^2 noté : $d_3(\sigma, P)$
- le test de χ^2 de NEYMAN'S noté : $d_4(\sigma, P)$
- le test basé sur l'information de SHANNON noté : $d_5(\sigma, P)$
- le test de WILCOXON noté : $d_6(\sigma, P)$
- le test utilisant la divergence de KULLBACK noté $d_7(\sigma, P)$.

Appliqués à la comparaison des deux lois P et R, ceux-ci s'écrivent respectivement (pour la démonstration voir annexe III.3).

a -

$$d_3(\sigma, P) = L_\alpha \frac{1 - \sum_1^\sigma P_{j/i}}{\sigma} = L_\alpha \frac{C_\sigma - 1}{C_\sigma + 1} = \chi^2 \quad (\text{III.17})$$

$$1 + \sum_1^\sigma P_{j/i}$$

χ^2 dépend de $(m-1)$ degrés de liberté. A partir d'un seuil ϵ , du χ^2 calculé et du χ^2 tabulé, il est alors possible d'accepter ou de rejeter l'hypothèse H_0 .

b -

$$d_4(\sigma, P) = \frac{1 - \sum_1^\sigma P_{j/i}}{\sum_1^\sigma P_{j/i}} = \frac{\sum_1^m P_{j/i}}{\sum_1^\sigma P_{j/i}} = C_\sigma - 1 = \chi^2 \quad (\text{III.18})$$

Nous remarquons que ce test conduit au même résultat que celui utilisant l'indice de distance basé sur le coefficient de contraste. L'interprétation d'un tel indice est identique au cas précédent.

c -

$$d_5(\sigma, P) = P_i \left[\left(1 - \frac{1}{C_\sigma} \right) \text{Log} \left(1 + \frac{1}{C_\sigma} \right) + \frac{1}{C_\sigma} \text{Log} \frac{1}{C_\sigma} \right] \quad (\text{III.19})$$

Cette quantité représente la perte d'information provoquée par l'opération de contraste. Elle est d'autant plus faible que C_σ est proche 1.

d -

$$d_6(\sigma, P) = \frac{1}{m\sigma - \frac{\sigma(\sigma+1)}{2}} - \frac{1}{U} \quad (\text{III.20})$$

où U est un coefficient qui tient compte de l'ordre des éléments des deux distributions après leur mélange (voir annexe III.3). Les résultats de ce test sont tabulés, il est alors possible d'accepter ou de rejeter l'hypothèse H_0 .

e -

$$d_7(\sigma, P) = -P_i \cdot \text{Log} \left(\sum_1^{\sigma} P_{j/i} \right) \quad (\text{III.21})$$

Un tel indice s'interprète également en termes de perte d'information comme l'indice d_5 .

En définitive, pour un niveau de contraste σ donné :

- . si $d(\sigma, P)$ est élevé : R ne se représente que faiblement la réalité du phénomène analysé, il est alors difficile de justifier son utilisation pour l'extraction des règles.
- . si $d(\sigma, P)$ est faible : les deux distributions sont "voisines", on peut alors envisager une extraction de règles à partir de R .

IV.3 Quelles connaissances peut-on extraire d'une ligne après l'opération de contraste ?

Nous avons vu que l'opération de contraste peut être acceptée ou rejetée selon le résultat du test. Ainsi :

- . si $\exists \sigma ; 1 \leq \sigma \leq m-1$ tel que H_0 est accepté : alors un contraste de niveau σ est effectué.
- . si $\nexists \sigma ; 1 \leq \sigma \leq m-1$ tel que H_0 est accepté : alors aucun contraste n'est opéré sur la ligne considérée, la seule valeur de σ pour laquelle H_0 est vraie est m .

Par conséquent à ce stade, on peut écrire, pour un niveau de contraste σ donné, avec H_0 vraie, une règle de type 1 ou type 2 :

$$\begin{aligned} \cdot X = \alpha_i \text{ ALORS } Y = \beta_1 \vee \beta_2 \vee \dots \vee \beta_\sigma & \quad \text{type 1} \\ \cdot X = \alpha_i \text{ ALORS } Y = (\beta_1, r_{1/i}) \vee (\beta_2, r_{2/i}) \vee \dots \vee (\beta_\sigma, r_{\sigma/i}) & \quad \text{type 2} \end{aligned}$$

On peut alors se poser la question de savoir, quel degré de vérité (ou de non vérité, quel degré d'efficacité il est possible d'attribuer à de telles règles ?

La réponse à cette question nous conduit à proposer certains coefficients de pondération qui vont dépendre essentiellement de la qualité de l'opération de contraste. Nous allons montrer que l'utilisation du test précédent ne conduit pas toujours à un "contraste optimal" et qu'il est nécessaire de faire appel à d'autres critères pour une évaluation correcte de celui-ci.

IV.4 Critères d'évaluation des seuils de contraste

IV.4.1 Position du problème

Le test effectué sur les deux distribution P et R, montre que la perte d'information, à l'issue de l'opération de contraste peut être faible ou élevée. Par ailleurs, quel que soit son niveau, cette information aurait pu ou non être exploitée. En effet, les données recueillies sur l'échantillon analysé peuvent se représenter de façons multiples. Soient les distributions suivantes (a, b, c, d) tracées sur la figure III.5.

a) Dans ce cas, la règle suivante peut être établie :

$$X = \alpha_i \text{ ALORS } Y = \beta_1 \vee \beta_2 \vee \beta_3 \vee \beta_4$$

la probabilité pour que cette règle soit vraie étant égale à 1. En effet, l'absence des résidus pour $\sigma > 4$ conduit, dans ce cas, à une perte d'information nulle et de ce fait l'information contenue dans la ligne est totalement exploitée par la règle ci-dessus.

b et c) Cette situation peut être modélisée par les deux règles suivantes :

$$\cdot X = \alpha_i \text{ ALORS } Y = \beta_1 \vee \beta_2 \vee \beta_3 \vee \beta_4 \dots (1)$$

$$\cdot X = \alpha_i \text{ ALORS } Y \notin \beta_1 \vee \beta_2 \vee \beta_3 \vee \beta_4 \dots (2)$$

avec la probabilité que la règle (1) soit vraie égale à 0.9 et la probabilité que la règle (2) soit vraie égale à 0.1.

. si la règle (2) existe dans la réalité, cela veut dire que la règle (1) ne représente pas complètement la réalité du phénomène analysé. Par contre si la règle (2) n'existe pas réellement, alors la règle (1) suffit à elle seule à modéliser totalement le phénomène étudié. Comment

peut-on distinguer ces deux cas en l'absence d'un modèle du système, la seule connaissance disponible étant les observations recueillies sur le système à travers les instruments de mesure. Ceux-ci peuvent en effet déformer la réalité (mauvaise acquisition, mesures aberrantes, comportements marginaux, ... etc.). Par ailleurs, les observations peuvent être également exprimées par un expert et, de la même façon, être entachées d'erreurs. Ceci nous conduit à étudier la répartition des résidus, les deux hypothèses extrêmes sont alors les suivantes :

- . H_0 : les résidus forment un seul élément (III;22)
- . H_1 : les résidus sont répartis de façon équiprobables

* Dans le cas où H_0 est réalisée, cela signifie que l'information résiduelle est faible et concentrée dans un seul élément dont la probabilité s'écrit :

$$P_{j/i} = P_{\sigma+1/i} = 1 - \sum_1^{\sigma} P_{j/i} \quad \text{avec} \quad P_{j/i} = 0 \quad \forall j \in \{\sigma+2, \dots, m\}$$

Par conséquent, la règle (1) n'est pas complète, il existe un phénomène non modélisé par celle-ci puisque l'information résiduelle peut être effectivement exploitable (cas b). Ainsi la règle (2) modélise la réalité qui n'a pas été prise en compte par la règle (1).

* Dans le cas où H_1 est réalisée, la répartition équiprobable des résidus peut être la conséquence d'une déformation de la réalité. L'information qu'ils contiennent est inexploitable, la probabilité de chaque élément s'écrit dans ce cas :

$$P_{j/i} = \frac{1 - \sum_1^{\sigma} P_{j/i}}{m - \sigma}, \quad \forall j \in \{\sigma+1, \dots, m\}$$

dans cet intervalle X et Y sont statistiquement indépendants, la connaissance de $X = \alpha_i$ n'apporte aucune information au sujet de Y. On peut alors estimer que la règle (1) est complète. L'inexploitabilité de l'information résiduelle ne justifie pas l'extraction de la règle (2).

d) Dans ce cas, on peut évidemment proposer la règle suivante :

$$X = \alpha_i \text{ ALORS } Y = \beta_1 \vee \beta_2 \vee \beta_3 \vee \beta_4$$

C'est une règle complète, puisqu'elle est proche de la réalité du phénomène analysé, mais elle est moins efficace pour l'utilisateur que la règle : $X = \alpha_i \text{ ALORS } Y = \beta_1 \vee \beta_2$ qui est également complète. En effet, l'information conservée (non résiduelle) contient une quantité d'information inexploitable qui a pour conséquence une diminution de l'efficacité de la règle ainsi générée.

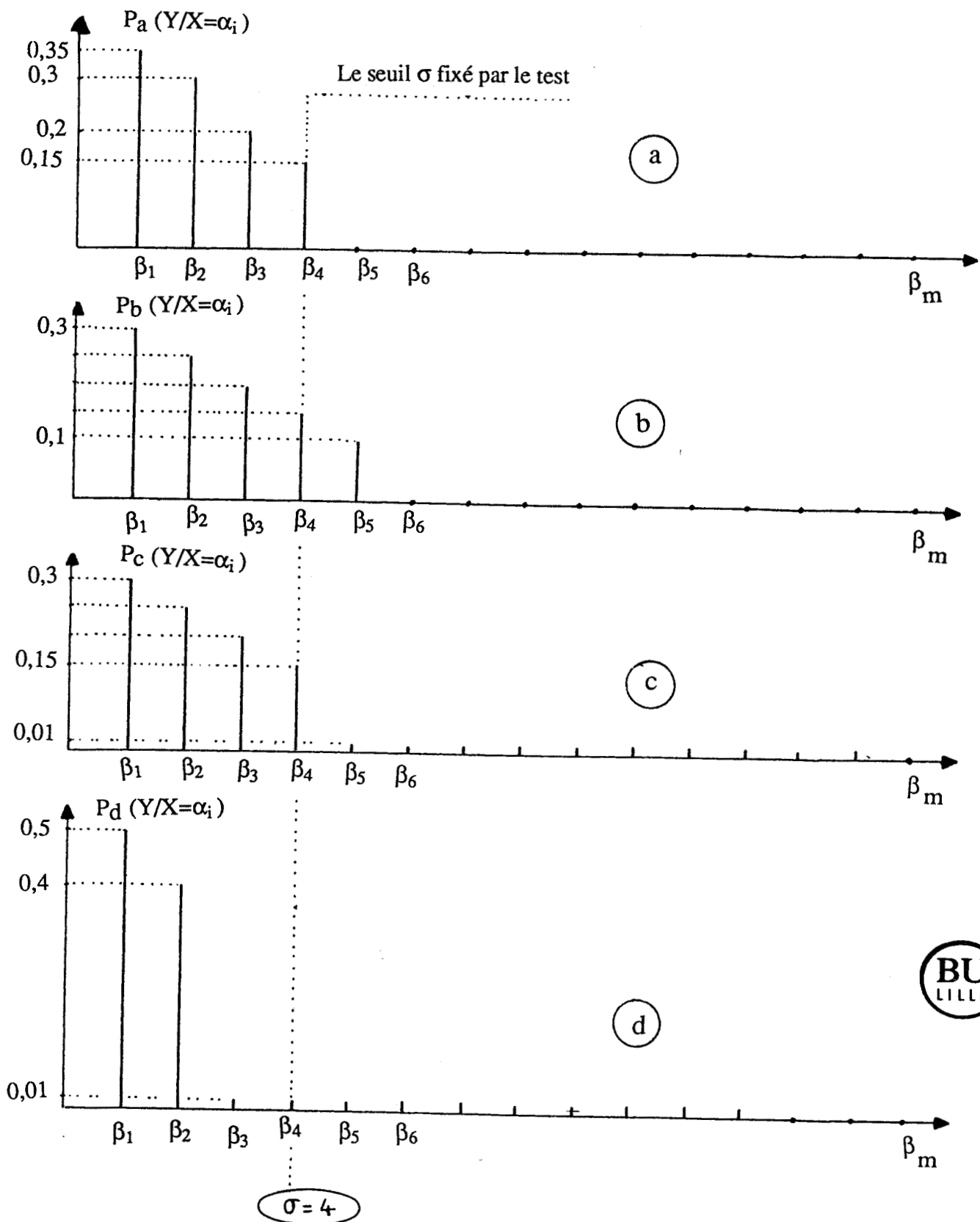


Fig III.5 tracé des distributions relatives aux cas : a, b, c et d

La détermination à partir du test précédent du seuil de contraste σ donnant naissance à des règles est valable dans le cas où les situations (a) et (c) sont réalisées. Cependant les deux autres cas peuvent justifier des contrastes plus importants (cas a) ou moins importants (cas b) suivant les possibilités d'utilisation de l'information résiduelle et la validité des règles obtenues. Pour tenir compte de ces éventualités, il est nécessaire de juger un seuil de contraste non seulement à partir du test précédent, qui juge la quantité d'information conservée (ou la conformité de la règle à la réalité observée) après contraste, mais aussi suivant d'autres critères basés sur les notions :

- de probabilité de la validité d'une règle ;
- de perte d'information ;
- d'efficacité d'une règle.

L'utilisation de ces trois critères, conjointement avec le test précédent conduit alors au choix d'un niveau de contraste "optimal" donnant naissance à des règles aussi proches que possible de la réalité du phénomène étudié.

IV.4.2 Jugement du contraste en termes de probabilité

Nous avons vu qu'à l'issue d'une opération de contraste de niveau σ , il est possible d'établir la règle suivante :

$X = \alpha_i$ ALORS $Y \in B_\sigma(\alpha_i)$; la probabilité pour que cette règle soit vraie est :

$$\sum_1^\sigma P_{j/i} \quad (\text{III.23})$$

et correspond au nombre de cas dans lesquels celle-ci a été observée vraie.

La probabilité pour que cette règle soit observée fautive est :

$$\left(1 - \sum_1^\sigma P_{j/i} \right)$$

et correspond à une règle complémentaire qui s'écrit :

$$X = \alpha_i \text{ ALORS } Y \notin B_\sigma(\alpha_i) \text{ OU } Y \in B(\alpha_i) \setminus B_\sigma(\alpha_i)$$

Autrement dit la probabilité pour qu'une règle soit observée vraie ou fautive dépend du niveau de contraste σ .

IV.4.3 Jugement du contraste en termes de perte d'information

L'opération de contraste de niveau σ , effectuée sur la distribution P , engendre une perte d'information mesurée par $d(\sigma, P)$. Or cette perte d'information, comme nous l'avons signalé au § 4.4.1., peut être significative ou non. En effet, celle-ci peut se présenter de façons multiples. Le problème qui se pose alors est celui de savoir dans quelle mesure la réalité observée peut elle se décomposer en un phénomène purement déterministe donnant naissance à une règle à laquelle se superpose un comportement purement aléatoire. La réponse à cette question nécessite l'étude de l'utilité de l'information résiduelle que nous noterons $U(\sigma)$ qui va être prise en compte pour juger le niveau de contraste σ . A cet effet, nous utilisons un modèle de comportement des résidus auquel va être comparée la distribution résiduelle réelle. Mais avant de présenter les différents modèles de comportement, nous allons rappeler dans un premier temps la distribution normalisée des résidus que l'on note Q et qui s'écrit :

$$Q(Y/X=\alpha) = \left\{ \begin{array}{l} q_{j/i}, j = \sigma+1, \dots, m \quad \text{tel que : } q_{j/i} = \frac{P_{j/i}}{\sum_{\sigma+1}^m P_{j/i}} \\ \text{avec } q_{j/i} = 0 \quad \text{pour } j = 1, \dots, \sigma \end{array} \right\}$$

IV.4.3.1 Le premier modèle

Soit $P(Y)$ la distribution moyenne de l'ensemble des lignes (distribution marginale) qui s'écrit : $P(Y) : \{P_j, j = 1, \dots, m\}$.

$$\text{Notons } S = \left\{ s_j, j = \sigma+1, \dots, m \quad \text{tel que : } s_j = \frac{P_j}{\sum_{\sigma+1}^m P_j} \right\} \quad (\text{III.24})$$

la distribution de probabilité supposée "résiduelle" obtenue à partir de $P(Y)$ sur l'intervalle $\{\sigma+1, \dots, m\}$. Le premier modèle consiste alors à comparer Q à S . Or S est issu de $P(Y)$ qui est le profil moyen de Y sur la population totale Ω et qui résulte de :

- . l'application simultanée à la population de toutes les règles ;
- . l'application d'un comportement purement aléatoire superposé aux règles.

Il n'existe aucune justification pour procéder de la sorte. En effet, comparer Q à S revient à faire le modèle d'aléa suivant :

- . il existe n règles, chacune d'elles est observée avec la probabilité P_i ;

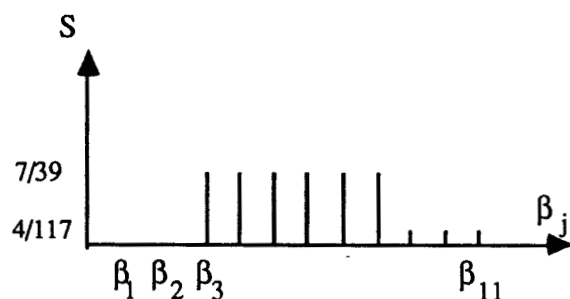
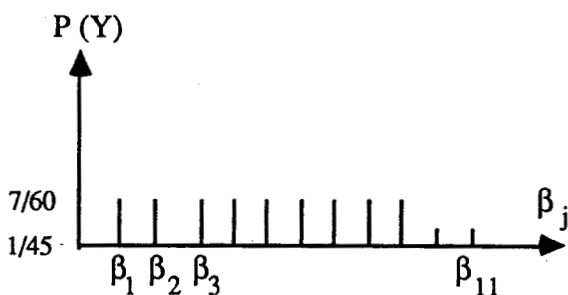
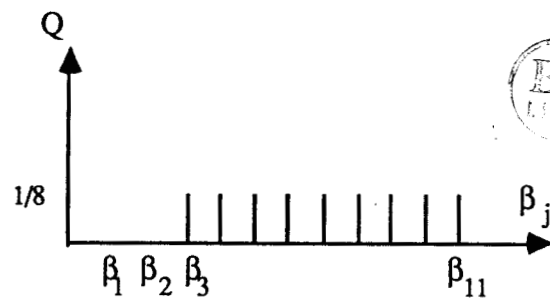
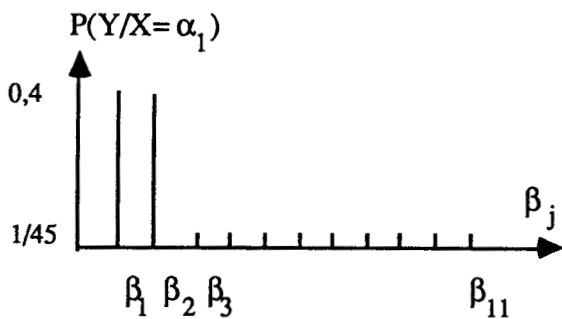
. il existe un comportement aléatoire qui consiste à appliquer l'une de ces n règles au hasard avec une distribution de probabilité P_i . Or, notre but est de chercher un modèle déterministe traduit par une règle auquel vient se superposer un comportement purement aléatoire. Pour confirmer l'inefficacité d'un tel modèle, prenons un exemple simple.

a) Exemple

Soit le tableau de contingence ci-dessous :

$X \backslash Y$	β_1	β_2	β_{11}										
α_1	0,1	0,1	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	0,25
α_2	$\frac{1}{180}$	$\frac{1}{180}$	0,1	0,1	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	0,25
α_3	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	0,1	0,1	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	0,25
α_4	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	0,1	0,1	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	$\frac{1}{180}$	0,25
	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{7}{60}$	$\frac{1}{45}$	$\frac{1}{45}$	$\frac{1}{45}$	

Traçons $P(Y/X = \alpha_1)$ et $P(Y)$ et comparons Q_1 à S .



Nous remarquons à travers cet exemple que les distributions $P(Y/X = \alpha_1)$ et $P(Y)$ sont totalement différentes. La première fait apparaître un phénomène déterministe que l'on peut modéliser par la règle : $X = \alpha_1$ ALORS $Y = \beta_1 \vee \beta_2$ auquel vient s'ajouter un phénomène aléatoire qui ne peut donner naissance à aucune règle. Par contre, la deuxième correspond à un profil moyen générant une règle qui est la résultante des quatre règles extraites des lignes et qui s'écrit :

$$X = \alpha_1 \vee \alpha_2 \vee \alpha_3 \vee \alpha_4 \text{ ALORS } Y = \beta_1 \vee \beta_2 \vee \dots \vee \beta_8$$

règle à laquelle se superpose également un phénomène non déterministe. Par conséquent rien ne justifie la comparaison de Q, qui est le résultat d'un comportement aléatoire, à S qui est en général la moyenne de différents comportements déterministes et aléatoires sur l'intervalle $\{\sigma+1, \dots, m\}$. En définitive, un tel modèle permet seulement d'apprécier de "combien" une règle issue d'une ligne s'écarte de la règle moyenne issue du profil moyen et ne peut être pris en compte pour juger l'utilité de l'information perdue. Par ailleurs, rappelons qu'en analyse factorielle des correspondances, la notion d'écartement du profil de chaque ligne (ou colonne) avec le profil moyen est mise en évidence. L'approche A.F.C. revient en réalité à mesurer de "combien" le profil d'une ligne s'écarte du profil moyen en utilisant la métrique du χ^2 centrée sur ce dernier. Par conséquent, l'A.F.C. permet d'étudier la dispersion des profils par rapport au profil moyen. Contrairement à l'A.F.C. notre approche consiste à chercher un modèle de comportement du système. En effet dans les cas limites où :

- ☛ tous les profils sont sensiblement voisins du profil moyen, le modèle du système peut être décrit par une seule et unique règle ;
- ☛ tous les profils sont différents, le modèle du système est décrit alors par les n règles atomiques.

IV.4.3.2 Les autres modèles de comportement des résidus

La recherche d'un modèle de comportement du système par élimination des modalités jugées sans signification suppose que l'on s'est fixé un modèle de comportement des résidus. En fait, décider si les résidus contiennent ou non une quantité d'information utile revient à comparer leur distribution conditionnelle de probabilité à la distribution des phénomènes aléatoires intervenus lors de la collecte des données (erreurs de mesure, bruits, ... etc). Celles-ci sont issues de "dispositifs" physiques (capteurs) ou humains (interviewés, experts), les aléas observés dépendant alors de la nature du capteur ainsi que de celle de la donnée. Le modèle d'aléa fournit alors les probabilités d'erreur supposées, sous la forme :

$$\text{Probabilité } [Y_{\text{observée}} = \beta_j / Y_{\text{réel}} \in B_{\sigma}(\alpha_i)], j \in \{\sigma + 1, \dots, m\} \quad (\text{III.25})$$

Selon la nature de la variable, du système, du capteur, ce modèle peut être élaboré de

différentes façons :

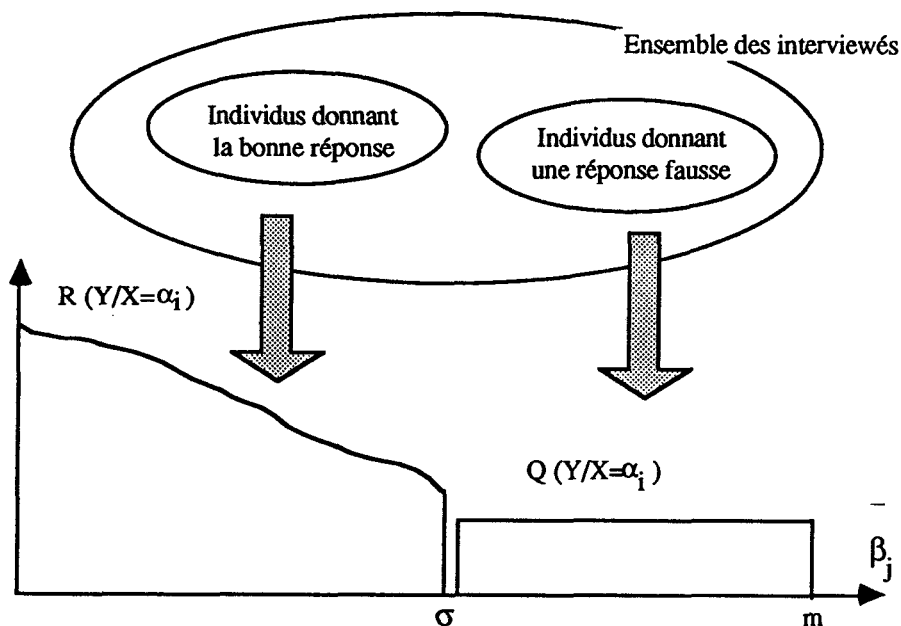
A - Ensemble de modalités non structurées

Dans le cas d'un ensemble de modalités non structurées et en absence totale d'information sur les probabilités d'erreur, la seule hypothèse que nous puissions faire est une hypothèse d'équirépartition des erreurs ou aléas. Dans ce sens, la distribution des résidus sera comparée à la distribution uniforme définie par :

$$Q(Y/X=\alpha_i) = \left\{ q_{j/i} , j = \sigma+1, \dots, m \quad \text{tel que : } q_{j/i} = \frac{\sum_{\sigma+1}^m P_{j/i}}{m - \sigma} \right\}$$

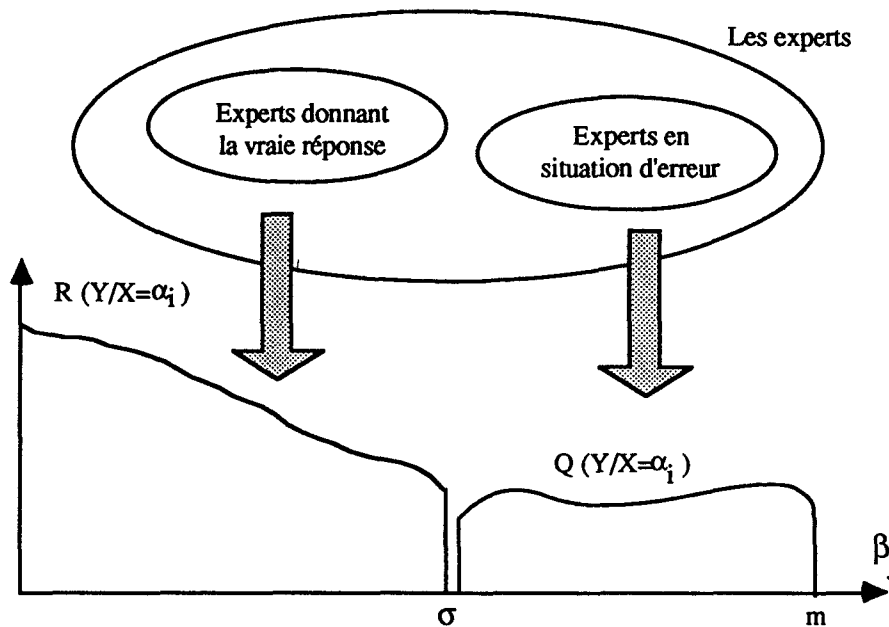
En effet, en absence d'information sur le système dans cet intervalle, rien ne justifie de privilégier une modalité par rapport à une autre.

Ce cas peut résulter d'un ensemble de réponses fournies par des individus à priori non spécialistes, les réponses fausses seront alors également réparties. Une telle situation peut être schématisée comme suit :



Cependant, lorsque "le capteur" utilisé est un expert, celui-ci peut non seulement fournir des informations relatives à la vérité du processus, mais aussi des informations relatives aux erreurs qu'il lui arrive de commettre, ou à celles commises le plus communément. Cette information sur les probabilités d'erreur peut être fournie directement par l'expert sous une forme

analogue à celle donnée par (III.25) ou encore par le biais d'une structuration de l'ensemble des modalités obtenue par la définition d'une relation floue, de ressemblance par exemple. A partir de la forme donnée par (III.25), on obtient directement le schéma suivant :



Dans le cas où l'expert fournit une structuration floue, il convient de bâtir à partir de celle-ci une distribution $Q(Y/X = \alpha_i)$ qui constitue le modèle des aléas. Ce passage de la structure sur M_Y (associée à la ligne $X = \alpha_i$) à la distribution $Q(Y/X = \alpha_i)$ est d'ailleurs à réaliser, que la structure soit floue et donnée par un expert ou stricte car existant a priori sur l'ensemble des modalités de Y .

B - Ensemble de modalités structurées

L'existence d'une structure sur l'ensemble des modalités de Y pourra toujours se traduire par la définition d'une distance (écart, dissemblance, dissimilarité, ...) entre une modalité "erronée" et l'ensemble des modalités vraies $B_\sigma(\alpha_i)$. Le passage à la distribution d'erreur $Q(Y/X = \alpha_i)$ utilisera cette distance par l'intermédiaire d'une loi de distribution donnée (par exemple exponentielle, normale, ... etc) sous la forme :

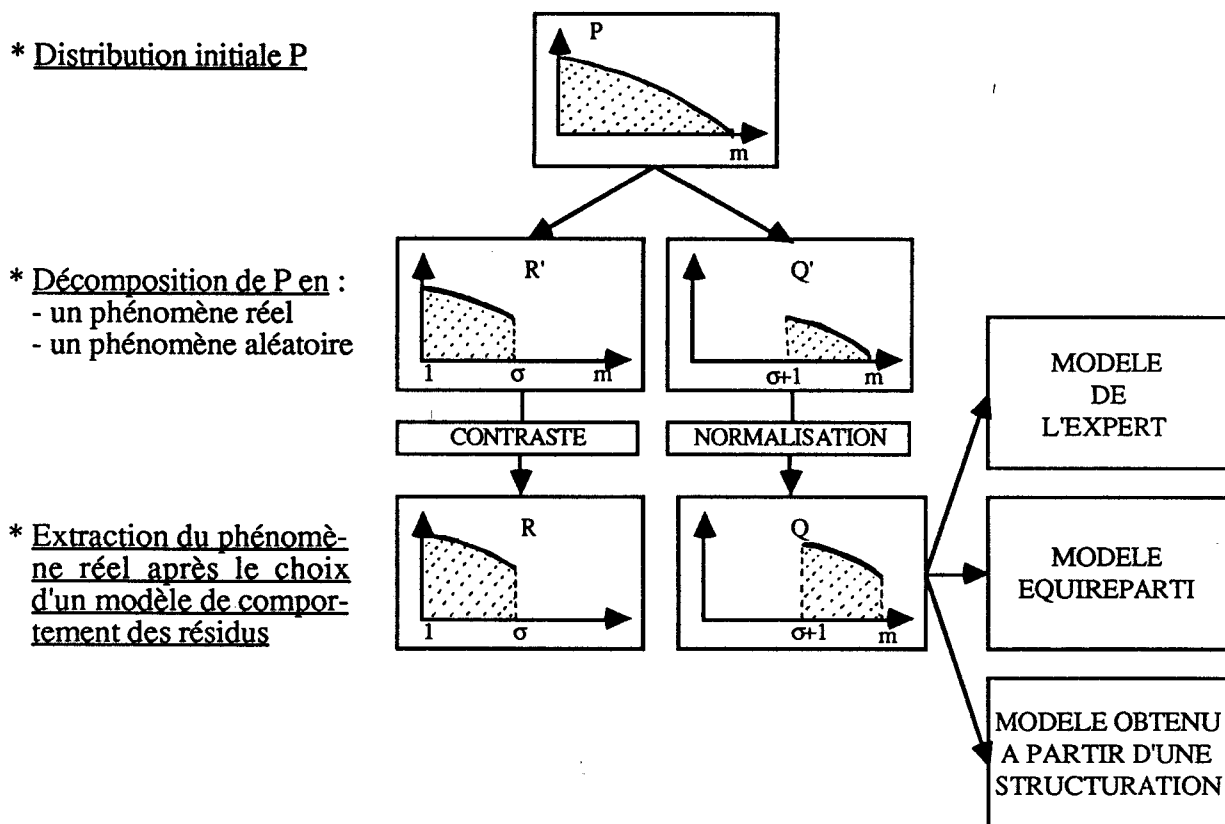
$$P [Y_{\text{observée}} = \beta_k / Y_{\text{réelle}} \in B_\sigma(\alpha_i)] = q [\delta(\beta_k, B_\sigma(\alpha_i))] \quad (\text{III.26})$$

où $\delta(\beta_k, B_\sigma(\alpha_i))$ désigne la distance entre β_k et $B_\sigma(\alpha_i)$ et q est une fonction décroissante de δ telle que :

$$q(\delta) \geq 0 \quad \forall \delta \geq 0$$

$$\sum_{k \in B_{\sigma}(\alpha_i)} q[\delta(\beta_k, B_{\sigma}(\alpha_i))] = 1$$

De façon schématique, la procédure adoptée peut se décomposer en trois niveaux que l'on peut résumer comme suit :



IV.4.3.3 Relation fondamentale



Les entropies $H(P)$, $H(Q)$, $H(R)$ des distributions initiale, résiduelle et contrastée sont liées par la relation suivante :

$$\left(\sum_1^{\sigma} P_{j/i} \right) H(R) + \left(\sum_{\sigma+1}^m P_{j/i} \right) H(Q) + H_2 = H(P) \quad (III.27)$$

avec H_2 l'entropie d'une variable à 2 classes "R et Q" qui s'écrit, si on note :

$$\sum_{\sigma+1}^m P_{j_i} = A$$

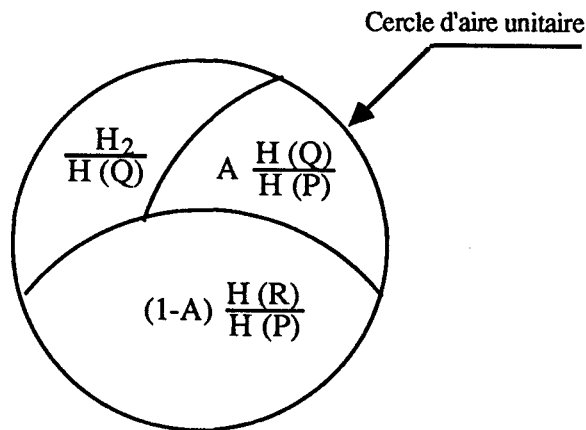
$$H_2 = - A \text{ Log } A - (1-A) \text{ Log } (1-A) \text{ avec } A \in [0,1]$$

Ainsi, la relation fondamentale (III.27) s'écrit sous-forme normalisée :

$$(1 - A) \frac{H(R)}{H(P)} + A \frac{H(Q)}{H(P)} + \frac{H_2}{H(P)} = 1 \quad (\text{III.28})$$

Cette relation est composée de 3 termes :

- . le 1^{er} est lié au pourcentage de l'information conservée ou fournie par l'opération de contraste ;
- . le 2^e est lié au pourcentage de l'information résiduelle perdue ;
- . le 3^e est lié au pourcentage de l'information qui ne peut être fournie sachant la règle de niveau σ .



Nous montrerons par la suite que les différents termes qui composent cette relation fondamentale vont être utilisés pour la définition des coefficients de vérité et d'efficacité d'une règle.

Pour définir l'utilité de l'information perdue $U(\sigma)$, considérons un modèle de bruit qui correspond à l'hypothèse d'une distribution uniforme de celui-ci.

IV.4.3.4 Définition de $U(\sigma)$

L'utilité de l'information perdue $U(\sigma)$ doit posséder les deux propriétés suivantes :

a) $U(\sigma)$ fonction strictement décroissante de $H(Q)$: (propriété P_1)

. si $H(Q) = H_M(Q)$: entropie maximale des résidus ; alors :

$U(\sigma) = U_m(\sigma)$: utilité minimale des résidus

En effet, la distribution Q s'écrit, en notant

$$\sum_{\sigma+1}^m P_{j/i} = A$$

$Q = \{q_{j/i}, j = \sigma+1, \dots, m \text{ tel que : } q_{j/i} = P_{j/i}/A \text{ et } q_{j/i} = 0 \text{ pour } j = 1, \dots, \sigma\}$

$H(Q)$ est maximale lorsque $P_{j/i} = A/(m-\sigma) \forall j \in \{\sigma+1, \dots, m\}$, autrement dit :

$$q_{j/i} = \frac{1}{A} \frac{A}{m-\sigma} = \frac{1}{m-\sigma} \quad (\text{équiprobabilité des résidus}) \text{ d'où :}$$

$$H_M(Q) = - \sum_{\sigma+1}^m \frac{1}{m-\sigma} \text{Log} \frac{1}{m-\sigma} = \text{Log}(m-\sigma) \quad (\text{III.29})$$

Par conséquent : $H(Q) = H_M(Q) = \text{Log}(m-\sigma)$ signifie que la règle de niveau σ est complète puisque les résidus ne contiennent aucune information susceptible d'être opérationnelle, elle n'est d'aucune utilité pour le phénomène modélisé par une telle règle.

. si $H(Q) = H_m(Q)$: entropie minimale des résidus, alors :

$U(\sigma) = U_M(\sigma)$: utilité maximale des résidus.

En effet, $H(Q)$ est minimale lorsque :

$$1 - \sum_{\sigma+1}^m P_{j/i} = 1 - A$$

c'est-à-dire $q_{\sigma+1/i} = A \cdot 1/A = 1$ (les résidus forment un seul élément) d'où :

$$H_m(Q) = 0 \quad (\text{III.30})$$

Autrement dit : $H(Q) = H_m(Q) = 0$ signifie que la règle de niveau σ est incomplète, les résidus contiennent encore une information susceptible d'être exploitable et qui n'a pas été prise en compte par une telle règle. Cette information résiduelle contient donc un phénomène non modélisé par la règle de niveau σ et celle-ci ne sera réellement complète qu'à l'issue de l'opéra-

tion de contraste de niveau $\sigma+1$.

A partir de ces deux cas limites, il semble raisonnable d'affecter une plus grande utilité à des résidus d'entropie faible. On aura alors :

$$H(Q_i) \leq H(Q_k) \Rightarrow U_i(\sigma) \geq U_k(\sigma) \quad (\text{III.31})$$

b) $U(\sigma)$ Fonction décroissante de σ : (propriété P_2)

L'utilité de l'information perdue doit être telle que :

$$U(\sigma) \geq U(\sigma+1) \quad (\text{III.32})$$

Ceci signifie que l'utilité de l'information perdue au niveau σ doit être supérieure ou égale à celle de l'information perdue au niveau $\sigma+1$ et ceci quelle que soit la distribution résiduelle. L'égalité étant obtenue lorsque les résidus sont répartis de façon équiprobable. Ceci semble logique puisque, d'une part, les raies de la distribution sont classées par ordre décroissant et d'autre part, si on juge au niveau σ que l'information résiduelle est inutile, il est difficile d'admettre qu'elle soit utile pour des niveaux de contraste plus élevés.

En notant

$$\sum_{\sigma+1}^m P_{j/i} = A$$

il semble logique de définir dans un premier temps l'utilité de l'information perdue, que nous noterons pour l'instant $U_1(\sigma)$, comme suit :

$$U_1(\sigma) = A \frac{[H_M(Q) - H(Q)]}{H_M(Q)} = A \left[1 - \frac{H(Q)}{\text{Log}(m - \sigma)} \right] \quad (\text{III.33})$$

Nous avons montré en annexe III.4 que l'utilité de l'information perdue ainsi définie vérifie la première propriété donnée par III.3.1. mais ne vérifie pas toujours la 2^e propriété (III.32). Par ailleurs, lorsque σ est voisin de m et particulièrement pour $\sigma = m-1$, $U_1(\sigma)$ n'est pas définie puisque : $H(Q) = H_M(Q) = 0$. Par conséquent $U_1(\sigma)$ ainsi définie, ne semble pas convenir pour évaluer l'utilité de l'information perdue dans le cas général. C'est pourquoi, une nouvelle définition de l'utilité de l'information perdue, s'avère nécessaire. A cet effet, nous proposons la fonction suivante que nous noterons $U_2(\sigma)$:

$$U_2(\sigma) = A \frac{[H_M(Q) - H(Q)]}{H_M(P)} \quad (\text{III.34})$$

$$U_2(\sigma) = A \frac{\text{Log}(m - \sigma) - H(Q)}{\text{Log } m} = \frac{\text{Log}(m - \sigma)}{\text{Log } m} \left[1 - \frac{H(Q)}{\text{Log}(m - \sigma)} \right]$$

$$\text{d'où } U_2(\sigma) = \frac{\text{Log}(m - \sigma)}{\text{Log } m} U_1(\sigma)$$

On montre en annexe III.4 que le fait d'avoir multiplié $U_1(\sigma)$ par la quantité $\text{Log}(m - \sigma) / \text{Log } m$, permet de lever le paradoxe énoncé ci-dessus. Par conséquent $U_2(\sigma)$ vérifie les deux propriétés de l'utilité de l'information perdue. En définitive, $U_2(\sigma)$ ainsi définie semble convenir à la définition d'une telle fonction dans le cas général et doit être prise également comme l'un des critères d'évaluation du seuil de contraste σ à retenir. Un tel seuil σ doit être choisi tel que l'utilité de l'information perdue reste minimale.

Quelques remarques :

- ☞ Les deux critères développés ci-dessus (en termes de probabilité, en termes de pertes d'information) vont être pris en compte pour juger de la vérité d'une règle dans un contexte plus global (cf. § IV.5.).
- ☞ $U_2(\sigma)$ que nous noterons désormais $U(\sigma)$ est homogène à une probabilité. Nous montrerons que $U(\sigma)$ n'est autre que le coefficient de non vérité de la règle extraite (cf. § IV.5.).

IV.4.4 Jugement en termes d'efficacité

L'efficacité d'une règle va dépendre essentiellement :

- du nombre σ de modalités prises par Y parmi les m modalités existantes (règles de type 1) ou
- de l'entropie de la distribution contrastée R (règles de type 2).

Ainsi une règle du type : $X = \alpha_i$ ALORS $Y = \beta_j$ est d'une grande efficacité pour l'utilisateur à cause de l'unicité de la conclusion. Par contre, une règle telle que :

$X = \alpha_i$ ALORS $Y = \beta_1 \vee \beta_2 \vee \dots \vee \beta_m$ n'est que d'une très faible efficacité pour l'utilisateur à cause de ses conclusions multiples. Cette efficacité est même nulle dans le cas des règles de type 1, puisque dans ce cas on n'a aucune information exploitable. Par ailleurs, le coefficient d'efficacité que nous noterons $e(\sigma)$ doit vérifier les propriétés suivantes :

- $e(1) = 1$
- $e(m) \geq 0$: pour une règle totalement inefficace on posera $e(m) = 0$ (III.35)
- $e(\sigma)$: fonction décroissante de σ

Nous allons envisager successivement le cas des règles de type 1 et de type 2.

a) Cas des règles de type 1

Dans cette situation, la probabilité de chacune des modalités β_j , $j = 1, \dots, \sigma$ n'est pas spécifiée. Celles-ci contribuent (vues sous l'angle de l'efficacité) de façon identique à l'efficacité de la règle. On leur attribue alors la même probabilité :

$$r_{j/i} = \frac{P_{j/i}}{\sum_1^{\sigma} P_{j/i}} = \frac{\sum_1^{\sigma} P_{j/i}}{\sigma \sum_1^{\sigma} P_{j/i}} = \frac{1}{\sigma}$$

L'entropie de la conclusion est alors égale à $\text{Log } \sigma$. Le même raisonnement appliqué à l'ensemble des modalités β_j , $j = 1, \dots, m$ conduit à une entropie égale à $\text{Log } m$.

Par ailleurs, en tenant compte des propriétés précédentes (III.35), la définition du coefficient d'efficacité (compris entre 0 et 1) est donné par :

$$e_1(\sigma) = \frac{\text{Log } m - \text{Log } \sigma}{\text{Log } m} = 1 - \frac{\text{Log } \sigma}{\text{Log } m} = \frac{\text{Log } \frac{m}{\sigma}}{\text{Log } m} \quad (\text{III.36})$$

- . si $\sigma = m \Rightarrow e_1(m) = 0$: la règle est totalement inefficace puisqu'elle n'apporte aucune information à l'utilisateur.
- . si $\sigma = 1 \Rightarrow e_1(1) = 1$: la règle est d'une efficacité maximale.

b) Cas des règles de type 2

Dans ce cas, la règle fait apparaître pour chaque modalité β_j sa probabilité $r_{j/i}$ et ceci quelque soit $j \in \{1, \dots, \sigma\}$. Par conséquent, l'efficacité de la règle doit tenir compte de la contribution de chaque modalité. Par ailleurs, même établie à partir des données non contrastées, une règle du type 2 peut avoir une certaine efficacité puisque $H(P) \leq \text{Log } m$. Dans ce sens, on définira l'efficacité d'une règle de type 2 bâtie à partir de la distribution R par :

$$e_2(\sigma) = \frac{\text{Log } m - H(R)}{\text{Log } m} = 1 - \frac{H(R)}{\text{Log } m} \quad (\text{III.37})$$

. si $\sigma = 1 \Rightarrow H(R) = 0 \Rightarrow e_2(1) = 1$: efficacité maximale

. si $\sigma = m \Rightarrow H(R) = H(P) \Rightarrow e_2(m) = \text{Log } m - H(P) / \text{Log } m$: efficacité initiale de la ligne

$e_2(m)$ apparaît alors comme une mesure de l'efficacité de l'information contenue dans une ligne du tableau initial.

Remarque IV.4.4.

$$e_2(\sigma) = 1 - \frac{H(R)}{\text{Log } m}$$

$$e_2(m) = 1 - \frac{H(P)}{\text{Log } m}$$

$$e_2(\sigma) - e_2(m) = \frac{H(P) - H(R)}{\text{Log } m} \Rightarrow e_2(\sigma) = e_2(m) + \frac{H(P) - H(R)}{\text{Log } m}$$

En notant $e_2(\sigma) = e_2(m) + \Delta e_2(\sigma)$, $\Delta e_2(\sigma)$ représente le gain d'efficacité obtenu par l'opération de contraste. Nous traiterons plus en détail ce problème au chapitre IV lors de la procédure de regroupement des lignes et les colonnes.

En définitive, les critères permettant de juger un contraste "optimal" et la règle qui en découle peuvent se résumer comme suit :

1 - Critère de conformité :

mesuré par $d(\sigma, P)$, celui-ci exprime la conformité de la règle à la réalité du phénomène observé.

2 - Probabilité d'occurrence :

mesurée par le modèle probabiliste

$$\sum_1^{\sigma} P_{j/i}$$

celui-ci fait apparaître le nombre de cas où une règle a été observée vraie.

3 - Critère de perte d'information :

mesuré par le coefficient $U(\sigma)$, celui-ci tient compte de l'utilité de l'information résiduelle et permet de conclure si la règle est complète ou non.

4 - Critère d'efficacité :

mesuré par $e_1(\sigma)$ (règles de type 1) ou $e_2(\sigma)$ (règles de type 2), celui-ci juge l'efficacité

d'une règle générée par un niveau de contraste donné.

Le choix du seuil de contraste σ doit être tel que :

$$\begin{aligned}
 & \cdot d(\sigma, P) \text{ minimum} \\
 & \cdot \sum_1^{\sigma} P_{j/i} \text{ maximum} \\
 & \cdot U(\sigma) \text{ minimum} \\
 & \cdot e(\sigma) \text{ maximum}
 \end{aligned}
 \tag{III.38}$$

Les trois premiers critères peuvent être utilisés pour définir le coefficient de vérité d'une règle. On peut remarquer que les deux premiers critères ne sont pas indépendants. En effet, pour les indices d_1 à d_7 , on observe une relation monotone décroissante entre $d(\sigma, P)$ et

$$\sum_1^{\sigma} P_{j/i}$$

Par conséquent, minimiser $d(\sigma, P)$ est donc équivalent à maximiser $\sum_1^{\sigma} P_{j/i}$

IV.5 Coefficients de vérité et de non vérité d'une règle

IV.5.1 Introduction

L'une des caractéristiques des systèmes experts est la manipulation de connaissances incertaines ou imprécises. La représentation et la formulation de ces connaissances, sous forme de règles est souvent déduite par un raisonnement approché qui traduit le savoir de l'expert dans le domaine analysé. Dans le cas de données incertaines, la notion de vérité ou de fausseté d'une proposition est introduite. Pour traduire cette incertitude, les experts introduisent soit directement, soit indirectement en s'aidant de la théorie Bayésienne des coefficients de vraisemblance [MARRAKCHI 1985], d'évocation/rejet [SOULA 1981], suffisance/nécessité [MARTIN 1984], crédibilité/plausibilité [SHAFER 1976], certitude [SHORTLIFE 1976] [DUDA 1980] ... etc. Par contre, dans le cas où les données sont imprécises, la théorie des sous-ensembles flous est alors exploitée pour tenir compte de l'interprétation de l'information dont on ne connaît qu'approximativement les limites [DUBOIS 1985]. Cette approche, basée sur l'utilisation des mesures floues de SUGENO [1974], sur des mesures de possibilité/nécessité [ZADEH 1978] [DUBOIS 1985], ... etc, s'est révélée adaptée à certains types de problèmes. Par ailleurs, très récemment, une analogie entre les intervalles de confiance utilisés en statistique et les intervalles de confiance que peut fournir un expert sur une proposition, a conduit G. PASS [1986] à formuler des règles de probabilité incertaine.

Dans le cas où le domaine analysé ne dispose pas d'experts ou bien lorsque ceux-ci éprouvent des difficultés à formuler des règles, la représentation des connaissances peut se faire

à l'aide des données statistiques recueillies sur celui-ci. Leur exploitation judicieuse peut donner naissance à des règles dont la validité est fonction de la représentativité de la population d'apprentissage. C'est pourquoi, nous allons proposer ci-dessous un coefficient de vérité basé sur les critères présentés aux paragraphes précédents.

IV.5.2 Coefficient de vérité d'une règle

Ce coefficient va être utilisé pour mesurer la vérité d'une règle. En notant $V(\sigma)$ le coefficient de vérité d'une règle issue d'un contraste de niveau σ , une approche probabiliste conduira à définir $V(\sigma)$ par :

$$V(\sigma) = \sum_1^{\sigma} P_{j/i} \quad (\text{III.39})$$

Cette approche présente plusieurs inconvénients [PRADE 1982]. Par ailleurs, dans sa formulation, cette approche ne permet pas de distinguer entre une règle complète qui exprime la vérité du phénomène analysé et une règle incomplète qui ne modélise que partiellement celui-ci. En effet, dans ce dernier cas, certains comportements du système ne sont pas modélisés par cette règle. Si ceux-ci sont peu fréquents, on peut décider en première approximation que la règle est vraie. C'est pourquoi, nous allons définir un coefficient de vérité qui tient compte non seulement du niveau de contraste σ , mais aussi de l'utilité de l'information résiduelle. Il semble raisonnable d'exiger d'un tel coefficient les trois propriétés suivantes :

- . $V(m) = 1$
- . $V(\sigma) = 1$ si $U(\sigma) = U_m(\sigma)$

$$V(\sigma) = \sum_1^{\sigma} P_{j/i} \quad \text{si } U(\sigma) = U_M(\sigma) \quad (\text{III.40})$$

Nous définissons le coefficient de vérité de niveau σ par : $V(\sigma) = 1 - U(\sigma)$ qui s'écrit encore :

$$V(\sigma) = \sum_1^{\sigma} P_{j/i} + \sum_{\sigma+1}^m P_{j/i} \left(1 - \frac{U(\sigma)}{\sum_{\sigma+1}^m P_{j/i}} \right) \quad (\text{III.41})$$

- . la quantité (1) n'est autre que le coefficient de vérité probabiliste.
- . la quantité (2) est un facteur correctif qui tient compte de l'utilité de l'information perdue. Il exprime également le fait que toute l'information utile de la ligne a été extraite ou non.

$V(\sigma)$ peut s'écrire également :

$$\begin{aligned}
V(\sigma) &= \sum_1^{\sigma} P_{j/i} + \sum_{\sigma+1}^m P_{j/i} \left(1 - \frac{H_M(Q) - H(Q)}{H_M(P)} \right) \\
&= 1 - \sum_{\sigma+1}^m P_{j/i} \frac{\text{Log}(m - \sigma) - H(Q)}{\text{Log } m}
\end{aligned}$$

. si $U(\sigma) = 0$ alors $v(\sigma) = 1$

$$\text{si } U(\sigma) \rightarrow \sum_{\sigma+1}^m P_{j/i} \quad \text{alors} \quad V(\sigma) \rightarrow \sum_1^{\sigma} P_{j/i} < 1$$

Bien que cela n'apparaisse pas explicitement dans la définition ci-dessus, l'existence d'une relation biunivoque entre $d(\sigma, P)$ et

$$\sum_1^{\sigma} P_{j/i}$$

(Cf § IV.2) montre que le coefficient de vérité défini ci-dessus fait intervenir la notion de conformité de la règle à la réalité. En s'aidant des relations existant entre $d(\sigma, P)$ et

$$\sum_1^{\sigma} P_{j/i}$$

le coefficient de vérité $V(\sigma)$ peut être exprimé par les relations suivantes :

- si $d(\sigma, P) = d_1(\sigma, P) = d_4(\sigma, P) : U(\sigma) = \sum_1^{\sigma} P_{j/i} [1 + d_1(\sigma, P)] - U(\sigma)$

- si $d(\sigma, P) = \frac{d_3(\sigma, P)}{L_{\alpha}} = \phi^2 : V(\sigma) = \sum_1^{\sigma} P_{j/i} [1 + d_3(\sigma, P)] + [d_3(\sigma, P) - U(\sigma)]$

- si $d(\sigma, P) = d_2(\sigma, P) : V(\sigma) = d_2(\sigma, P) + \frac{H(R)}{H(P)} - U(\sigma)$

Remarque 1 : Nous constatons que $V(\sigma)$ est indépendant de $e(\sigma)$. En effet, il est tout à fait possible qu'une règle vraie soit inefficace (un truisme).

Remarque 2 : Comme nous avons défini $V(\sigma)$, il est alors possible de définir le coefficient de non vérité $NV(\sigma)$ d'une règle par :

$$NV(\sigma) = 1 - V(\sigma) = U(\sigma) = \sum_{\sigma+1}^m P_{j/i} \left[\frac{\text{Log}(m - \sigma) - H(Q)}{\text{Log } m} \right] \quad (\text{III.42})$$

Le coefficient de non vérité d'une règle n'est autre que le coefficient de l'utilité de l'in-

formation perdue comme nous l'avons signalé au § IV.4.3. En effet, si l'information résiduelle est exploitable, elle est donc utile, par conséquent, il est possible d'extraire une règle vraie sur l'intervalle $\{\sigma+1, \dots, m\}$. Par contre, si les résidus ne recèlent aucune information opérationnelle, leur utilité est donc nulle et ils ne peuvent donner naissance à une règle vraie.

V - EXTRACTION DE CONNAISSANCES : APPROCHE ANALYSE STRUCTURALE PAR LA THEORIE DE L'INFORMATION

V.1 Introduction

Comme l'approche précédente, cette approche est également locale et consiste par conséquent à extraire des règles des lignes et des colonnes du tableau de contingence. elle utilise les concepts de la théorie de l'information et adopte une procédure pas à pas basée sur la définition d'un coefficient d'extraction de l'information qui dépend entre autres de la répartition de l'information résiduelle.

Si on se réfère à la relation I.56 (§ 4.3.2. chapitre I), l'entropie conditionnelle de Y sachant X, s'écrit :

$$H(Y/X) = \sum_i P(X=\alpha_i) \cdot H(Y/X=\alpha_i)$$

où $H(Y/X=\alpha_i)$ est l'entropie de Y relative à la ligne i du tableau $[P_{ij}]$: "quantité d'information locale". Que vaut $H(Y/X)$ respectivement dans le cas d'un système déterministe et dans le cas d'un système non déterministe ?

a) Cas d'un système déterministe :

Pour un tel système $H(Y/X) = 0$, ce qui signifie que le tableau $[P_{ij}]$ contient un seul élément par ligne. Nous avons montré, au chapitre II, § II.2.1., que cette dernière égalité implique qu'à partir de chaque ligne i, $i=1, \dots, n$, il est possible d'extraire la règle suivante :

$$X = \alpha_i \text{ ALORS } Y = \beta_j \text{ avec } V = P_{j/i} = 1 \quad \text{---} \quad \text{(III.43)}$$

Dans le cas où deux lignes i et k aboutissent à la même conclusion, leur regroupement conduit à la règle suivante :

$$X = \alpha_i \vee \alpha_k \text{ ALORS } Y = \beta_j \text{ avec } V = P_{j/i \vee k} = 1 \quad \text{---} \quad \text{(III.44)}$$

Dans le cas particulier où :

$H(Y/X) = H(X/Y) = 0$ (système inversible), il s'agit juste d'un recodage.

b) Cas d'un système non déterministe :

Dans un système complexe, il est rare que les tableaux de contingence contiennent un seul élément par ligne. Les mesures sont généralement bruitées et entachées d'erreur et en plus des aléas physiques et/ou technologiques, les cases de $[P_{ij}]$ sont souvent non nulles et forment une distribution dont les éléments sont la résultante d'un phénomène réel pouvant donner naissance à une règle à laquelle vient se superposer un phénomène purement aléatoire dû à la distorsion de la réalité. Dans une telle situation, la génération des règles devient délicate, et il s'agit alors de définir les critères visant à faciliter cette extraction. Ceci nous conduit naturellement, dans un premier temps, à étudier la répartition des résidus.

V.2 Répartition des résidus

. Considérons une ligne i du tableau de contingence $[P_{ij}]$, $i \in \{1, \dots, n\}$ et soit $P(Y/X = \alpha_i) = \{P_{j/i}, j=1, \dots, m\}$ sa distribution conditionnelle de probabilité.

- . Soit $P_{j/i} = \max P_{k/i}, k=1, \dots, m$
- . Appelons pour l'instant résidus : $\{P_{k/i}, k \neq j, j=1, \dots, m\}$

L'entropie conditionnelle de Y sachant $X = \alpha_i$ s'écrit dans ce cas :

$$H(Y/X=\alpha_i) = -P_{j/i} \text{Log} P_{j/i} - \sum_{k \neq j} P_{k/i} \text{Log} P_{k/i} \quad (\text{III.45})$$

Calculons le majorant et le minorant de $H(Y/X = \alpha_i)$ sous la contrainte :

$$\sum_{k \neq j} P_{k/i} = 1 - P_{j/i} \quad (\text{III.46})$$

. Le majorant de $H(Y/X = \alpha_i)$

$$\begin{aligned} &\text{Calculons } \max \left[- \sum_{k \neq j} P_{k/i} \text{Log} P_{k/i} \right] \\ &\text{sous la contrainte } \sum_{k \neq j} P_{k/i} = 1 - P_{j/i} \end{aligned}$$

La solution est évidente puisque $-\sum P_{k/i} \text{Log} P_{k/i}$ est maximum lorsque les $P_{k/i}$ sont répartis de façon équiprobable et égaux à

$\frac{1 - P_{j/i}}{m - 1}$. Dans ce cas $P(Y/X = \alpha_i)$ s'écrit :

$$P(Y/X = \alpha_i) = \left\{ P_{j/i}, \frac{1 - P_{j/i}}{m - 1}, \dots, \frac{1 - P_{j/i}}{m - 1} \right\} \quad \text{d'où une entropie maximale}$$

$$H_M = H_M(Y/X = \alpha_i) = -P_{j/i} \text{Log } P_{j/i} - (1 - P_{j/i}) \text{Log } \frac{1 - P_{j/i}}{m - 1} \geq H(Y/X = \alpha_i) \quad (\text{III.47})$$

. Le minorant de $H(Y/X = \alpha_i)$

$$\text{Calculons } \min \left[- \sum_{k \neq j} P_{k/i} \text{Log } P_{k/i} \right]$$

$$\text{sous la contrainte } \sum_{k \neq j} P_{k/i} = 1 - P_{j/i}$$

Le minimum s'obtient lorsque tous les $P_{k/i}$ sont nuls sauf un dont la probabilité vaut : $1 - P_{j/i}$ d'où une entropie minimale :

$$H_m = H_m(Y/X = \alpha_i) = -P_{j/i} \text{Log } P_{j/i} - (1 - P_{j/i}) \text{Log } (1 - P_{j/i}) \leq H(Y/X = \alpha_i) \quad (\text{III.48})$$

Par conséquent : $H_m(Y/X = \alpha_i) = H_m \leq H(Y/X = \alpha_i) = H \leq H_M(Y/X = \alpha_i) = H_M$

La connaissance de H et les deux quantités extrêmes H_m et H_M va nous permettre de définir deux coefficients que nous noterons η_m et η_M , pour juger l'état de répartition des résidus. Ils sont définis comme suit :

$$\eta_M = \frac{H_M - H}{H_M - H_m} \quad (\text{III.49})$$

$$\eta_m = \frac{H - H_m}{H_M - H_m} \quad (\text{III.50})$$

On remarque que $\eta_M + \eta_m = 1$

- . si $H \rightarrow H_M$ alors $\eta_M \rightarrow 0$ et $\eta_m \rightarrow 1$: les résidus ont tendance à se répartir de façon équiprobable.
- . si $H \rightarrow H_m$ alors $\eta_M \rightarrow 1$ et $\eta_m \rightarrow 0$: les résidus ont plutôt tendance à se regrouper et former un seul élément.

Pour mieux apprécier l'état de répartition des résidus, nous allons tracer H_M , H_m et H en fonction de $P_{j/i}$ à $\eta = \text{constante}$.

a) $\eta = \eta_M$:

$$H = H_m + (1 - \eta_M) (1 - P_{j/i}) \text{Log} (m-1) = H_M - \eta_M (1 - P_{j/i}) \text{Log} (m-1)$$

$$\frac{\partial H_M}{\partial P_{j/i}} \text{ s'annule pour } P_{j/i} = \frac{1}{m}$$

$$\frac{\partial H}{\partial P_{j/i}} \text{ s'annule pour } P_{j/i} = \frac{1}{1 + 2^k} \quad \text{avec } k = (1 - \eta_M) \text{Log} (m - 1) = \text{Cte}$$

on montre facilement que $\frac{1}{1 + 2^k} \geq \frac{1}{m}$, en effet,

$$(1 - \eta_M) \text{Log} (m-1) \leq \text{Log} (m-1) \text{ puisque } : 0 \leq \eta_M \leq 1$$

$$(m - 1)^{(1 - \eta_M)} \leq (m - 1) \Rightarrow \frac{1}{1 + (m - 1)^{(1 - \eta_M)}} \geq \frac{1}{m} \quad \text{d'où :}$$

$$\frac{1}{1 + 2^k} \geq \frac{1}{m}$$

b) $\eta = \eta_m$

$$H = H_m + \eta_m (1 - P_{j/i}) \text{Log} (m-1) = H_M + (\eta_m - 1) (1 - P_{j/i}) \text{Log} (m-1)$$

$$\frac{\partial H_M}{\partial P_{j/i}} \text{ s'annule pour } P_{j/i} = \frac{1}{2}$$

$$\frac{\partial H}{\partial P_{j/i}} \text{ s'annule pour } P_{j/i} = \frac{1}{1 + 2^{k'}} \quad \text{avec } k' = \eta_m \text{Log} (m - 1) = \text{Cte}$$

on montre facilement que $\frac{1}{1 + 2^k} \leq \frac{1}{2}$ puisque $2^k \geq 1$ pour $m > 2$

$$\text{par conséquent } \frac{1}{m} \leq \frac{1}{1 + 2^k} \leq \frac{1}{2}$$

Traçons ci-dessous (fig. III.6) H , H_M et H_m en fonction de $P_{j/i}$ à $\eta = \eta_m = \eta_M = 1/2$

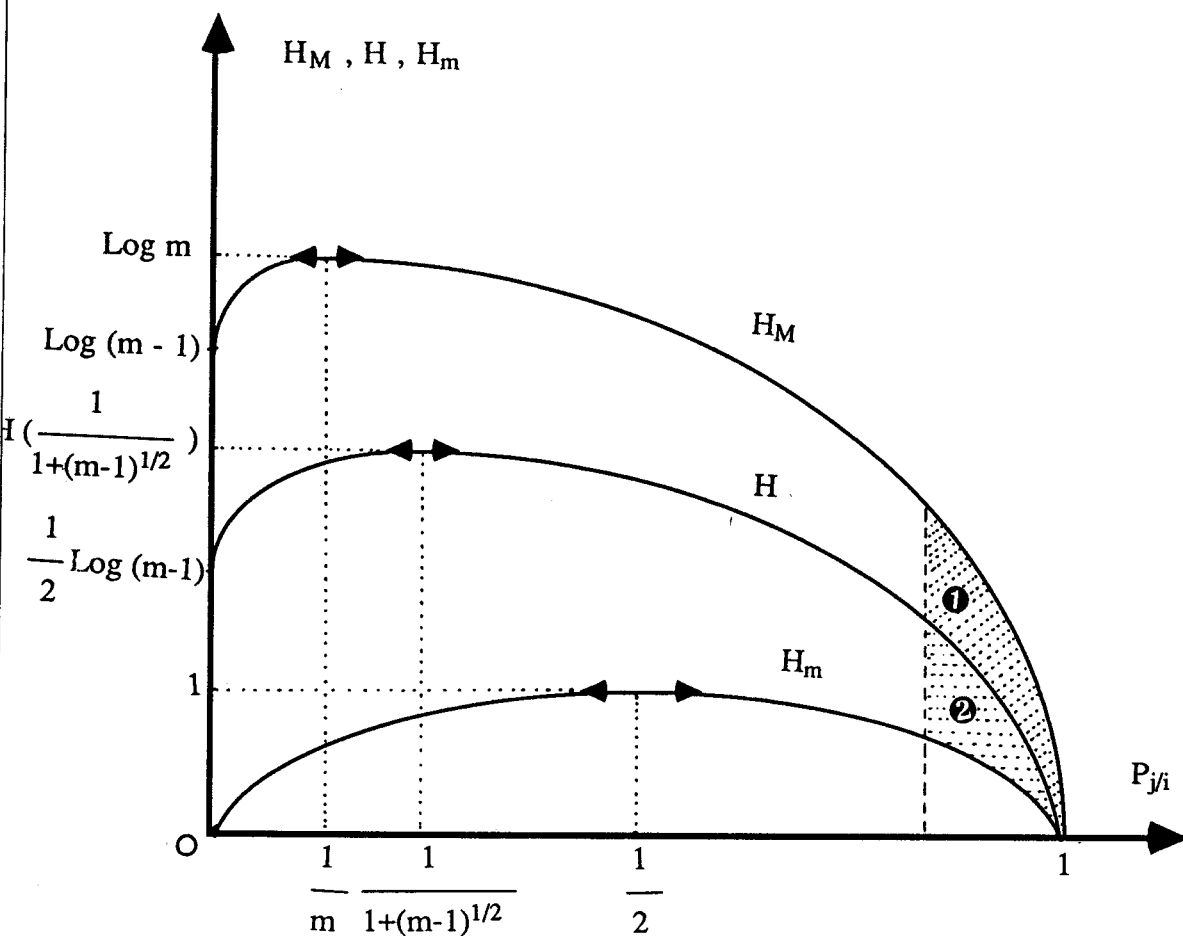


Fig. III.6 : tracé de H_M , H , H_m en fonction de $P_{j/i}$ à $\eta_M = \eta_m = 1/2$

- ❶ Dans cette zone, les résidus ont tendance à se répartir de façon équiprobable.
- ❷ Dans cette zone, les résidus ont tendance à se regrouper en un seul élément.

La connaissance de l'état de répartition des résidus va nous guider pour extraire l'information contenue dans une ligne du tableau de contingence et évaluer la qualité. Son exploitation nous permettra de définir un coefficient d'extraction de l'information.

V.3 Exploitation de l'information contenue dans une ligne

L'exploitation du contenu informationnel d'une ligne va conduire à l'extraction d'une règle. Pour ce faire, nous allons dans un premier temps ordonner les éléments de la distribution ligne par ordre décroissant, c'est-à-dire :

$$P(Y/X = \alpha_i) = \{P_{j/i}, j = 1, \dots, m \text{ tel que : } P_{1/i} \geq P_{2/i} \geq \dots \geq P_{m/i}\}$$

Appelons Y_σ la variable Y après réunion de ses σ premières modalités β_1 à β_σ , avec :

$1 \leq \sigma \leq m$. L'entropie d'une telle variable s'écrit alors :

$$H(Y_\sigma / X = \alpha_i) = - \left(\sum_1^\sigma P_{j/i} \right) \text{Log} \left(\sum_1^\sigma P_{j/i} \right) - \sum_{\sigma+1}^m P_{j/i} \text{Log} P_{j/i} = H_\sigma \quad (\text{III.51})$$

- . si $\sigma = 1 \Rightarrow H(Y_1/X = \alpha_i) = H(Y/X = \alpha_i)$: aucun regroupement des modalités n'est opéré.
- . si $\sigma = m \Rightarrow H(Y_m/X = \alpha_i) = 0$: toutes les modalités de Y son regroupées en une seule.

On effectue un tel regroupement dans le but de déceler les modalités les plus significatives et les plus importantes de la ligne considérée. Autrement dit, cette façon de procéder nous conduit à déceler les "raies" les plus influentes du spectre de la distribution $P(Y/X = \alpha_i)$. Ce qui revient en définitive à se poser la question suivante : dans quelle mesure le regroupement de niveau σ conduisant à la règle $X = \alpha_i$ ALORS $Y = \beta_1 \vee \dots \vee \beta_\sigma$ est-il tel que l'on puisse affirmer que les $(m-\sigma)$ modalités restantes ne sont dûes qu'à un comportement aléatoire et que cette règle est conforme à la réalité observée. En termes de partitions, ceci revient à générer une partition de $Y/X = \alpha_i$ en $(m-\sigma+1)$ classes à partir de sa partition atomique à m classes. La résolution d'un tel problème nécessite la définition d'un coefficient d'extraction, puisque le calcul de H_σ définie par la relation (III.51) ne suffit pas à lui seul à conclure sur la validité et la conformité de la règle précédente à la réalité. En effet, si on appelle résidus : $\{P_{j/i}, j = \sigma+1, \dots, m\}$, H_σ peut être faible ou élevée et comme nous l'avons signalé au paragraphe consacré à l'utilité de l'information perdue, il ne peut distinguer d'une manière opérationnelle, entre information résiduelle exploitable et information résiduelle non exploitable. D'où la nécessité de définir un coefficient d'extraction de l'information. Mais avant d'aborder sa formulation, nous allons décomposer H_σ en deux parties : information conservée et information résiduelle et voir comment elles évoluent. Soit :

$$H_\sigma = - \left(\sum_1^\sigma P_{j/i} \right) \text{Log} \left(\sum_1^\sigma P_{j/i} \right) - \sum_{\sigma+1}^m P_{j/i} \text{Log} P_{j/i} = h_0 + h \quad (\text{III.52})$$

$$H_M = - \left(\sum_1^\sigma P_{j/i} \right) \text{Log} \left(\sum_1^\sigma P_{j/i} \right) - \left(1 - \sum_1^\sigma P_{j/i} \right) \text{Log} \left(\frac{1 - \sum_1^\sigma P_{j/i}}{m - \sigma} \right) = h_0 + h_M$$

$$H_m = - \left(\sum_1^\sigma P_{j/i} \right) \text{Log} \left(\sum_1^\sigma P_{j/i} \right) - \left(1 - \sum_1^\sigma P_{j/i} \right) \text{Log} \left(1 - \sum_1^\sigma P_{j/i} \right) = h_0 + h_m$$

H_M, H_m étant respectivement les entropies maximale et minimale de $Y/X = \alpha_i$ après réunion de ses σ premières modalités.

Traçons h, h_M et h_m en fonction de h_0 (fig. III.7) et (fig. III.8).

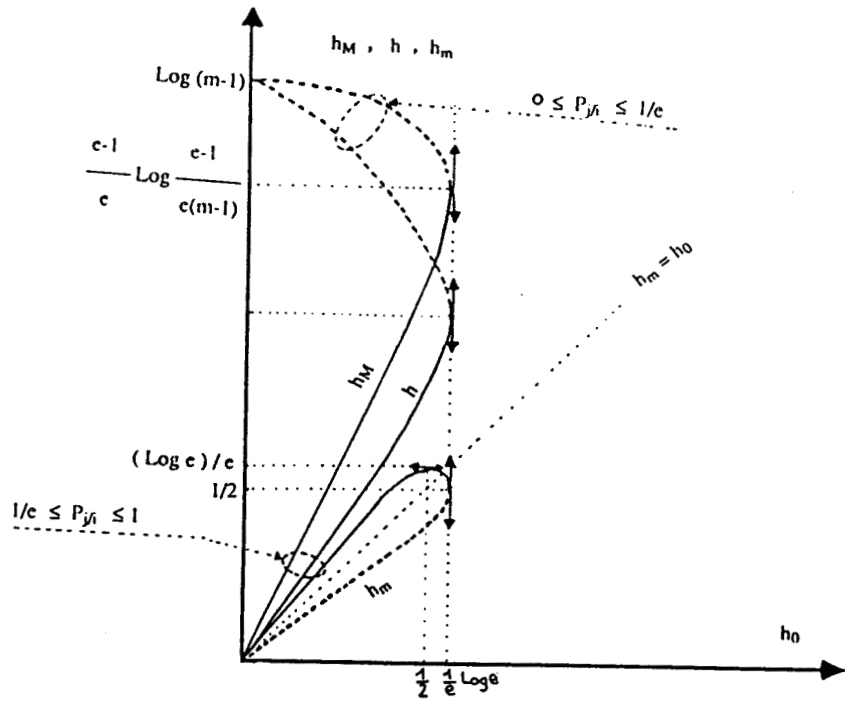


Fig. III. 7 Tracé de h, h_M et h_m en fonction de h_0

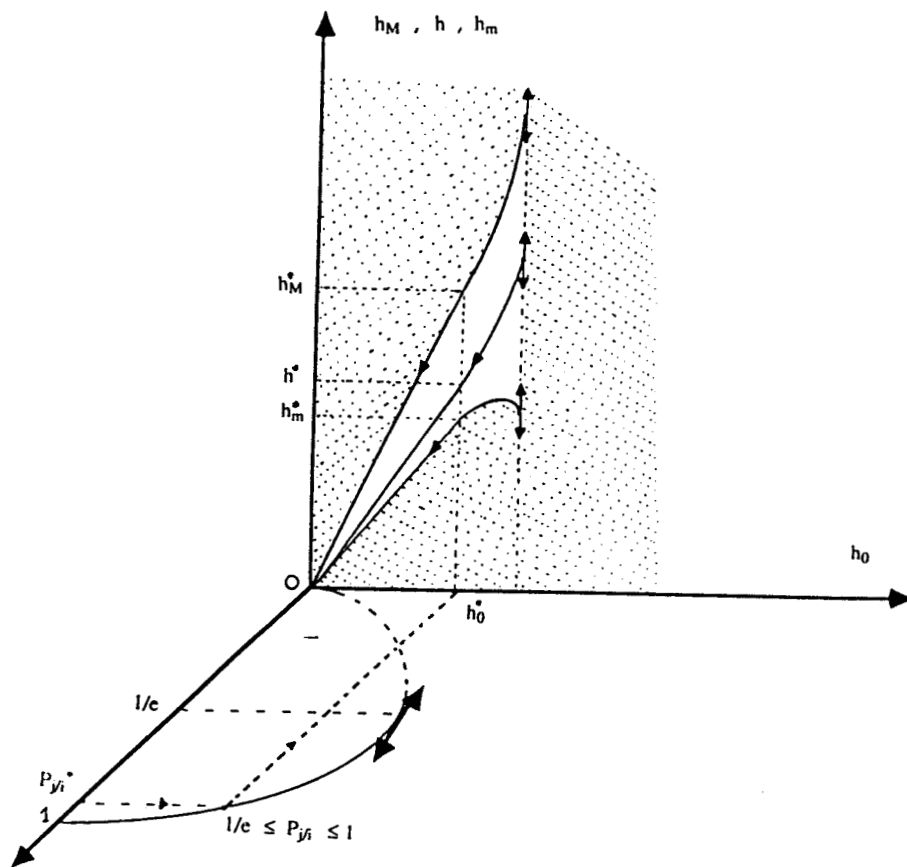


Fig III.8: Zone de variation de h en fonction de h_0 (partie non hachurée)



- ☞ On constate que lorsque $1/2 \leq P_{j/i} \leq 1$ la courbe $h = f(h_0)$ (fig. III.6) est majorée par $h_M = f(h_0)$ et minorée par $h_m = g(h_0)$.
- ☞ A h_0 constant, h varie entre h_M et h_m
- ☞ Nous nous sommes intéressés au tracé de h en fonction de h_0 pour $1/e \leq P_{j/i} \leq 1$ (fig. III.7) puisque au niveau σ , $P_{j/i}$ est supérieur à $1/e$ et c'est dans cet intervalle que l'on juge l'information résiduelle. A l'aide du formalisme dont on dispose à ce niveau, il est possible de définir des coefficients permettant de juger la qualité d'une règle.

V.4 Coefficient d'extraction de l'information

V.4.1 Définitions

Le regroupement des modalités de Y engendre une perte d'information. Cette perte d'information est proportionnelle à la diminution d'incertitude sur Y . En effet, si on note :

- . H : l'information totale initiale
- . H_σ : l'information non encore utilisée au niveau σ
- . $H_{\sigma+1}$: l'information non encore utilisée au niveau $\sigma+1$

. $(H - H_\sigma)$ et $(H - H_{\sigma+1})$: représentent respectivement la perte d'information engendrée par σ , $\sigma+1$ regroupements ; c'est aussi [l'information totale initiale] - [l'information non encore utilisée au niveau σ , $\sigma+1$]

. H_σ peut être :

- . faible et utile : c'est le cas où H_σ est au voisinage de H_m
- . élevée mais inutile : c'est le cas où H_σ est au voisinage de H_M .

C'est pourquoi, le coefficient d'extraction de l'information va dépendre de H_σ et $H_{\sigma+1}$, nous proposons alors :

$$\begin{aligned}
 \xi(\sigma) &= (H - H_{\sigma+1}) - (H - H_\sigma) = H_\sigma - H_{\sigma+1} && \text{(III.53)} \\
 &= [\text{l'information utilisée au niveau } \sigma+1] - [\text{information utilisée au niveau } \sigma] \\
 &= [\text{réduction de l'incertitude au niveau } \sigma+1] - [\text{réduction de l'incertitude au niveau } \sigma] \\
 &= [\text{information susceptible d'être extraite entre } \sigma \text{ et } \sigma+1] : \text{ c'est pourquoi, nous avons} \\
 &\quad \text{appelé ce coefficient : le } \underline{\text{coefficient d'extraction de l'information}}.
 \end{aligned}$$

$$\xi(\sigma) = H_\sigma - H_{\sigma+1}$$

- ☞ Si cette quantité est non négligeable, cela veut dire qu'il est encore possible de réduire l'incertitude sur Y , autrement dit la ligne contient encore une information utile.

☛ Si cette quantité est négligeable, cela signifie qu'on ne peut plus réduire l'incertitude sur Y, il est alors inutile d'effectuer d'autres regroupements puisque l'information susceptible d'être extraite par cette opération serait inexploitable et diminue inutilement l'efficacité de la règle.

Afin de juger correctement la variation $H_\sigma - H_{\sigma+1}$, nous allons normaliser cette quantité compte tenu du fait que H est la seule grandeur qui reste constante lors de l'opération de regroupement, le coefficient d'extraction de l'information normalisé est défini par :

$$E(\sigma) = \frac{H_\sigma - H_{\sigma+1}}{H} = \frac{\xi(\sigma)}{H} \quad (\text{III.54})$$

Si on note ϵ un seuil d'extraction d'information limite, alors :

- . $E(\sigma) \leq \epsilon$: signifie que l'information qui reste à extraire n'est pas opérationnelle.
- . $E(\sigma) > \epsilon$: signifie que l'information qui reste à extraire peut être opérationnelle.

V.4.2 Etude de $E(\sigma)$:

Ecrivons de façon explicite $E(\sigma)$:

$$E(\sigma) = \frac{H_\sigma - H_{\sigma+1}}{H} = \frac{1}{H} \left(\sum_1^{\sigma+1} P_{j/i} \right) \left[- \frac{P_{\sigma+1/i}}{\sum_1^{\sigma+1} P_{j/i}} \text{Log} \frac{P_{\sigma+1/i}}{\sum_1^{\sigma+1} P_{j/i}} - \frac{\sum_1^\sigma P_{j/i}}{\sum_1^{\sigma+1} P_{j/i}} \text{Log} \frac{\sum_1^\sigma P_{j/i}}{\sum_1^{\sigma+1} P_{j/i}} \right]$$

La quantité entre crochets n'est autre que l'entropie d'une variable à deux modalités ayant respectivement comme probabilité :

$$\left(\frac{P_{\sigma+1/i}}{\sum_1^{\sigma+1} P_{j/i}} \right) \quad \text{et} \quad \left(1 - \frac{P_{\sigma+1/i}}{\sum_1^{\sigma+1} P_{j/i}} \right) = \frac{\sum_1^\sigma P_{j/i}}{\sum_1^{\sigma+1} P_{j/i}}$$

Par ailleurs, si on note : $\sum_1^\sigma P_{j/i} = P$ et $P_{\sigma+1/i} = \alpha P$, avec $0 \leq \alpha \leq 1$

$$E(\sigma) \text{ s'écrit alors : } E(\sigma) = \frac{P}{H} [(1+\alpha) \text{Log}(1+\alpha) - \alpha \text{Log} \alpha] \quad (\text{III.55})$$

Traçons $E(\sigma)$ en fonction de α (fig. III.9) avec P et H constants :

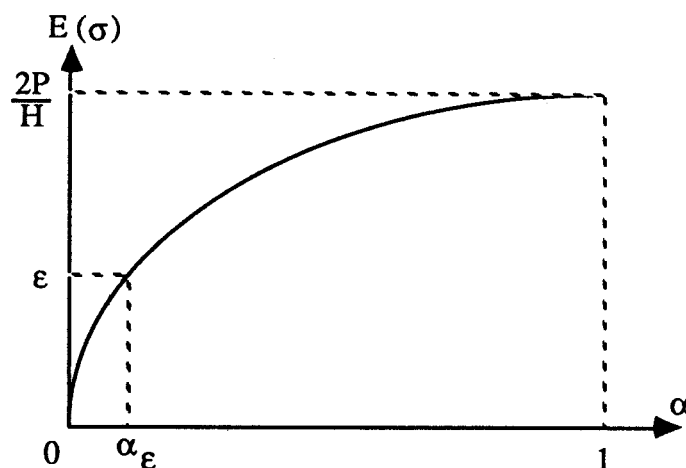


Fig III.9 : tracé de $E(\sigma)$ en fonction de α

Le choix d'un seuil d'extraction limite ϵ , conduit à conclure si l'information qui reste à extraire est opérationnelle ou non. En effet :

$$\text{si } P_{\sigma+1/i} \text{ est faible devant } \sum_1^{\sigma} P_{j/i} \Rightarrow E(\sigma) \leq \epsilon \Rightarrow \alpha \leq \alpha_{\epsilon}$$

$$\text{si } P_{\sigma+1/i} \text{ est non négligeable devant } \sum_1^{\sigma} P_{j/i} \Rightarrow E(\sigma) > \epsilon \Rightarrow \alpha > \alpha_{\epsilon}$$

Cas particulier :

$$\text{si } \alpha = 1 \Rightarrow P_{\sigma+1/i} = P = \sum_1^{\sigma} P_{j/i}$$

$$\text{or } P_{\sigma/i} \geq P_{\sigma+1/i} \text{ et } \sum_1^m P_{j/i} = 1 \Rightarrow m=2 \text{ et } P_{\sigma/i} = P_{\sigma+1/i} = 0,5$$

$$\Rightarrow H_2 = 0 \text{ et } H_1 = H = 1 \text{ et } E(\sigma) = E(1) = 1$$

A ce stade, le problème qui se pose est celui de savoir à quel niveau le processus d'extraction de l'information doit s'arrêter. Comme pour le cas du niveau de contraste, le niveau d'extraction limite σ est jugé non seulement par $E(\sigma)$ qui fait apparaître la diminution de l'entropie (entre σ et $\sigma+1$), mais aussi par d'autres critères faisant intervenir les notions :

- . d'utilité de l'information non extraite
- . de probabilité d'occurrence

- . de réduction d'information
- . d'efficacité de la règle générée.

V.5 Les critères de jugement du niveau d'extraction limite

V.5.1 Jugement en terme d'utilité de l'information non extraite

Si au niveau σ , l'information non utilisée n'est pas opérationnelle, alors la règle qui en découle sera qualifiée de complète. Dans le cas contraire, elle sera jugée incomplète et le seuil d'extraction limite à retenir se trouve à un niveau supérieur à σ . Celui-ci va dépendre effectivement de l'utilité de l'information non utilisée que nous noterons $U(\sigma)$.

Si on se réfère aux relations (III.49), (III.50) et (III.52), on peut écrire alors :

$$\eta_M(\sigma) = \frac{H_M - H_\sigma}{H_M - H_m} = \frac{h_M - h}{h_M - h}$$

et

$$\eta_m(\sigma) = \frac{H_\sigma - H_m}{H_M - H_m} = \frac{h - h_m}{h_M - h_m}$$

(III.56)

L'utilisation de ces deux quantités permet de définir l'utilité de l'information non utilisée par :

$$U(\sigma) = \left(\sum_{\sigma+1}^m P_{j/i} \right) (1 - \eta_m(\sigma)) = \left(\sum_{\sigma+1}^m P_{j/i} \right) \eta_M(\sigma)$$

$$= \left(\sum_{\sigma+1}^m P_{j/i} \right) \frac{H_M - H_\sigma}{H_M - H_m}$$

(III.57)

Si on note $Q = \left\{ q_{j/i}, j = \sigma+1, \dots, m \text{ tel que } q_{j/i} = \frac{P_{j/i}}{\sum_1^m P_{j/i}} \text{ et } q_{j/i} = 0 \text{ pour } j = 1, \dots, \sigma \right\}$

la distribution normalisée des résidus, et $A = \sum_{\sigma+1}^m P_{j/i}$

alors $H(Q)$, H_σ et H_m sont liés par la relation suivante :

$$H_{\sigma} - H_m = A \cdot H(Q) \quad (\text{III.58})$$

or l'utilisation des relations (III.58) et (III.56) nous conduit à écrire $U(\sigma)$ comme suit :

$$U(\sigma) = A \cdot \left[1 - \frac{H(Q)}{H_M(Q)} \right] \quad \text{avec} \quad H_M(Q) = \text{Log}(m - \sigma)$$

qui n'est autre que l'utilité de l'information perdue $U_1(\sigma)$ définie dans l'approche précédente (cf relation III.33).

Or $U_1(\sigma)$ ainsi définie vérifie la première propriété de l'utilité de l'information perdue (cf § IV.4.3.5.a) mais ne vérifie pas toujours sa deuxième propriété (cf § IV.4.3.5.b). C'est pourquoi, et pour les mêmes raisons évoquées dans l'approche précédente, il semble raisonnable de définir l'utilité de l'information non encore utilisée par :

$$U(\sigma) = \frac{H_M - H_{\sigma}}{\text{Log } m} = A \left(\frac{H_M(Q) - H(Q)}{\text{Log } m} \right) \quad (\text{III.59})$$

V.5.2 Jugement en termes de probabilité d'occurrence :

(identique à l'approche précédente)

V.5.3 Jugement en termes de réduction d'information

Celle-ci est mesurée par un coefficient que nous noterons $r(\sigma)$ et qui fait apparaître la quantité d'information non encore utilisée H_{σ} au profit de la règle à générer. On le définit sous forme normalisée par :

$$r(\sigma) = H_{\sigma} / H(P) \quad (\text{III.60})$$

qui représente le pourcentage de l'information non encore utilisée au niveau σ . Plus $r(\sigma)$ est faible, plus l'information qui reste à extraire est négligeable, mais cela ne veut pas dire qu'elle est inutile. Par conséquent, ce coefficient doit être pris en compte pour juger le niveau d'extraction limite σ .

V.5.4 Jugement en termes d'efficacité

Comme dans l'approche précédente, nous distinguerons l'efficacité des règles de type 1 et de type 2.

a) Les règles de type 1 : (l'efficacité est identique à l'approche précédente)

b) Les règles de type 2 :

En faisant un raisonnement analogue à celui de l'approche précédente, on définit l'efficacité de cette règle par :

$$e_2(\sigma) = 1 - \frac{H - H_\sigma}{\text{Log } m} \quad (\text{III.61})$$

Or, dans ce cas, les règles font apparaître pour chaque modalité $\beta_j, j = 1, \dots, \sigma$ un coefficient de pondération probabiliste $r_{j/i}$ avec $\sum r_{j/i} = 1$. On obtient par conséquent une nouvelle distribution de probabilité R qui n'est autre que la distribution P normalisée ou contrastée et de ce fait :

H devient $H(R)$ et $H_\sigma = 0$ d'où

$$e_2(\sigma) = 1 - \frac{H - H_\sigma}{\text{Log } m} = 1 - \frac{H(R)}{\text{Log } m}$$

On obtient, donc par cette approche le même coefficient d'efficacité que l'approche précédente.

V.6 Coefficient de vérité

Celui-ci est défini par :

$$V(\sigma) = 1 - U(\sigma)$$

or $U(\sigma)$ est défini de façon identique par les deux approches, par conséquent $V(\sigma)$ peut s'écrire alors :

$$V(\sigma) = \sum_1^\sigma P_{j/i} + \sum_{\sigma+1}^m P_{j/i} \left(1 - \frac{\text{Log}(m - \sigma)}{\text{Log } m} \eta_M(\sigma) \right) \quad (\text{III.62})$$

Le coefficient de non vérité de la règle est défini par :

$$\begin{aligned} NV(\sigma) = U(\sigma) &= \left(\sum_{\sigma+1}^m P_{j/i} \right) \frac{h_M - h}{h_M - h_m} \\ &= \left(\sum_{\sigma+1}^m P_{j/i} \right) \frac{\text{Log}(m - \sigma)}{\text{Log } m} \eta_M(\sigma) \end{aligned} \quad (\text{III.63})$$

On peut remarquer également que $V(\sigma) + NV(\sigma) = 1$

En définitive le choix du seuil d'extraction limite "optimal" doit être tel que :

$$\left. \begin{array}{l} E(\sigma) \text{ minimum (coefficient d'extraction)} \\ r(\sigma) \text{ minimum (critère de réduction d'information)} \\ U(\sigma) \text{ minimum (critère de perte d'information)} \\ \sum_1^{\sigma} P_{j/i} \text{ maximum (critère probabiliste)} \\ e(\sigma) \text{ maximum} \end{array} \right\} \Rightarrow V(\sigma) \text{ maximum}$$

Nous remarquons que les deux premiers critères ne sont pas indépendants. En effet,

$$E(\sigma) = (H_{\sigma} - H_{\sigma+1}) / H = H_{\sigma} / H - H_{\sigma+1} / H = r(\sigma) - r(\sigma+1)$$

V.7 Comparaison des deux approches

Bien que les formulations des deux approches soient différentes, elles permettent d'une manière opérationnelle de générer les mêmes règles avec les mêmes coefficients de vérité et d'efficacité si le niveau de constance et le seuil d'extraction sont identiques dans les deux cas. Ceci apparaît clairement lorsque on se réfère aux deux relations fondamentales données respectivement par (III.27) et (III.58) et que nous écrivons ci-dessous :

$$\begin{array}{l} \cdot (1 - A) H(R) + A H(Q) + H_2 = H(P) \\ \cdot H_{\sigma} - H_m = A H(Q) \end{array}$$

En remplaçant les différentes entropies par leurs valeurs respectives, on montre que ces deux relations sont identiques puisque :

$$\begin{array}{l} \cdot H_2 = H_m \text{ (entropie d'une variable binaire)} \\ \text{et } \cdot H_{\sigma} = H(P) - (1-A) H(R) \end{array} \quad \text{--- (III.64)}$$

avec

$$A = \sum_{\sigma+1}^m P_{j/i}$$

$H(P)$ étant constante, le choix du seuil σ à retenir consiste, dans la première approche,

à raisonner en termes de contraste [H(R)] alors que dans la deuxième approche, on fait apparaître la notion de réduction d'information (H_σ).

Finalement dans les deux cas, la détermination du niveau σ "optimal" pour ligne i , $i = 1, \dots, n$ et pour chaque colonne j , $j = 1, \dots, m$ conduit à la génération de $(n+m)$ règles. Chacune d'elles est munie d'un coefficient de vérité et d'un coefficient d'efficacité et constituent ensemble le modèle redondant du système.

VI - COMPARAISON DES DEUX APPROCHES AVEC L'APPROCHE ANALYSE DES DONNEES

VI.1 Introduction

Nous avons signalé au chapitre I, que l'une des méthodes les plus puissantes de l'analyse des données est l'analyse factorielle des correspondances (A.F.C.) L'objectif principal d'une telle méthode est d'aboutir à des résultats clairement interprétables. Ainsi M. VOLLE [1976] dit à ce propos "une méthode d'analyse des données, n'est pas un objet logique dont on aurait tout dit une fois qu'on l'aurait décrit logiquement ; c'est un outil destiné à être utilisé conjointement avec d'autres pour classer, découper, visualiser, résumer, bref, décrire des gisements de données que leur volume rendrait inutilisables autrement". Dans l'approche système expert, plusieurs auteurs ont fait appel à l'analyse des données. Ainsi V. RIALLE [1985] utilise l'analyse en composantes principales pour proposer un système d'aide à la formulation de connaissances expertes appliqué à la génétique médicale. Dans ce sens, E. DEMONCHAUX [1985] propose également un système expert en analyse des données (CLAVECIN) pour réaliser un environnement d'assistance intelligent offrant ainsi le choix entre plusieurs méthodes d'analyse.

C'est pourquoi, nous allons utiliser l'A.F.C. comme un outil supplémentaire en vue de compléter et de générer les connaissances sur le système analysé.

VI.2 Analyse des tableaux contrastés

. L'ensemble des lignes fournissent un tableau contrasté dont la distribution conjointe de chaque ligne s'écrit, $\forall i, i = 1, \dots, n$:

$$R_{\sigma_i} = \left\{ r_{ij} = P_{ij} \frac{1}{C_{\sigma_i}} \quad , \quad j = 1, \dots, \sigma_i \quad \text{et} \quad r_{ij} = 0 \quad \text{pour} \quad j = \sigma_i + 1, \dots, m \right\}$$

. De même, l'ensemble des colonnes du tableau donnent naissance à un tableau contrasté dont la distribution conjointe de chaque colonne s'écrit :

$\forall j, j = 1, \dots, m$

$$R_{\sigma_j} = \left\{ t_{ij} = P_{ij} \frac{1}{C_{\sigma_j}} \quad , \quad i = 1, \dots, \sigma_j \quad \text{et} \quad t_{ij} = 0 \quad \text{pour} \quad i = \sigma_j + 1, \dots, n \right\}$$

A ce niveau, il est possible de générer à partir du tableau initial $[P_{ij}]$ les deux tableaux suivants :

$[r_{ij}]$: tableau de contingence contrasté sur les lignes

$[t_{ij}]$: tableau de contingence contrasté sur les colonnes

L'application de l'A.F.C. sur l'ensemble de ces deux tableaux, permet de fournir une représentation et une interprétation graphique de l'ensemble des règles qui constituent le modèle du système. Une telle représentation est alors plus proche de la réalité que celle que peut fournir le tableau initial $[P_{ij}]$ qui analyse les données brutes qui sont susceptibles de contenir une quantité d'information non opérationnelle. Les connaissances extraites de l'ensemble de ces tableaux sont à comparer avec l'A.F.C. appliquée à un tableau de contingence approximé. En effet, l'exploitation des résultats de l'A.F.C. devient intéressante lorsqu'il est possible d'expliquer une grande partie de l'inertie du nuage de points représentatifs du système, à partir d'un sous-espace p de faible dimension. Ce qui conduit naturellement à envisager l'analyse du tableau $[P_{ij}]$ approximé noté $[\phi_{ij}]^P$.

VI.3 Analyse de $[\phi_{ij}]^P$

Nous donnons en annexe III.5, les notations et quelques rappels des résultats de l'A.F.C. obtenus à partir de l'analyse de $[P_{ij}]$ et $[P_{ij}]$ approximé noté $[\phi_{ij}]^P$. Le problème qui se pose pour l'exploitation de tels tableaux est celui du choix du niveau d'approximation P de sorte que l'information perdue à l'issue de cette opération reste minimale. Ainsi au moyen d'un test du χ^2 il est possible de comparer les lignes et les colonnes de $[P_{ij}]$ avec celles de $[\phi_{ij}]^P$. Un tel test permet de juger, pour un niveau d'approximation p donné, si les 2 lignes (respectivement les 2 colonnes) sont approximativement identiques ou différentes. Celui-ci est connu en général sous le nom d'instruction "écart" et permet d'identifier les points pour lesquels la perte d'information occasionnée par la projection sur le sous-espace des p premiers vecteurs propres est la plus grande. Une telle perte d'information est donnée par exemple lorsque la décomposition est effectuée selon les lignes, par :

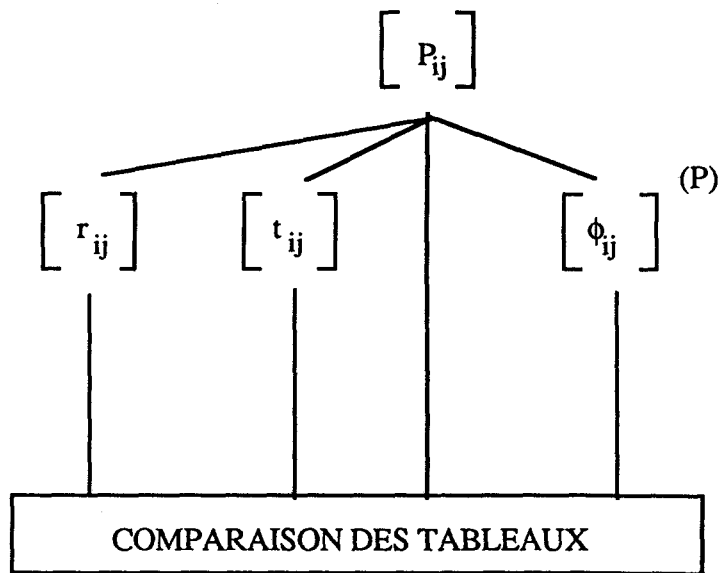
$$\sum_i \frac{(P_{ij} - \phi_{ij}^P)^2}{P_i \cdot P_j} \quad (III.65)$$

Elle permet d'identifier les éléments qui sont à l'origine de cette perte. En revanche, elle

ne peut distinguer d'une manière opérationnelle entre l'utilité ou la non-utilité de l'information perdue.

VI.4 Récapitulation

Si l'objectif visé par les deux approches présentées est la recherche d'un modèle sous formes de règles, l'utilisation de l'A.F.C. constitue un moyen supplémentaire de jugement des connaissances extraites du tableau de contingence. Une étude comparative des résultats issus de l'application de l'A.F.C. sur les différents tableaux énoncés ci-dessus permettra, d'accélérer et de faciliter l'extraction des règles, donc le modèle du système. D'une façon schématique une telle comparaison peut s'effectuer comme suit :



VII - CONCLUSION

Dans ce travail, nous avons présenté deux approches permettant l'extraction des connaissances à partir d'un tableau de contingence. La première approche utilise un opérateur de contraste, la seconde fait appel à une procédure d'extraction pas à pas, elles conduisent ensemble sous certaines conditions à la même représentation des connaissances. Celles-ci sont présentées sous forme de règles, chacune d'elles est dotée d'un coefficient de vérité et d'un coefficient d'efficacité, elles constituent ensemble le modèle du système. L'utilisation de ces approches conjointement avec l'analyse factorielle des correspondances peut faciliter l'expression de ces connaissances en l'absence du modèle du système. Les règles qui en découlent constituent alors le modèle provisoire du système. En effet, il convient de s'assurer que certaines règles ne constituent pas différentes écritures d'une seule et unique règle. Dans ce cas, une procédure de regroupement des règles est alors nécessaire et conduit au modèle simplifié du système qui peut

être modifié ou mis à jour à chaque fois que de nouvelles données sont disponibles. Une telle procédure fera l'objet du chapitre suivant.

D'une manière générale l'approche proposée constitue un extracteur de connaissance pouvant être utilisé en modélisation, en reconnaissance des formes, mise à jour en vérification des règles d'un système expert. Une telle approche est particulièrement bien adaptée aux différents domaines du diagnostic : diagnostic médical, diagnostic des pannes, détection des modes de fonctionnements (normal, défaillant, ...) ... etc. On développera à cet effet, au chapitre IV, une application dans le domaine d'aide à la décision médicale.

CHAPITRE IV

SIMPLIFICATION DU MODELE OBTENU
PAR L'APPROCHE SYSTEME EXPERT

I - INTRODUCTION

Nous avons montré au chapitre précédent qu'il est possible de générer $(n+m)$ règles à partir des partitions atomiques de X et de Y (partitions les plus fines ou partitions discrètes). Les règles ainsi extraites constituent le modèle redondant du système, qu'il convient de simplifier. En effet, dans le cas où n et m sont élevés et/ou certaines règles aboutissent à des conclusions identiques ou voisines, il est possible de procéder à leur regroupement, c'est-à-dire d'agréger les lignes et les colonnes. Cette façon de procéder conduit, lorsque cette opération est possible, à des règles composées, ce qui permet après élimination des règles redondantes, une diminution notable de la taille de la base des règles. L'ensemble de ces règles non redondantes forme le modèle simplifié du système qui peut être modifié ou mis à jour à chaque fois que de nouvelles données sont disponibles. Regrouper les lignes, par exemple, revient à rechercher des partitions moins fines que la partition atomique de l'ensemble des modalités de X : $P_n(M_X) = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$. Ce problème se présente et peut être formulé comme un problème de décomposition ou de classification. Il s'agit en effet, de chercher pour chaque niveau k , $1 \leq k \leq n$, une partition de X en k classes $P_k(X)$ telle que la perte d'information occasionnée par cette opération soit la plus faible possible et conduise à une structure simplifiée du modèle. Ceci, nous amène naturellement à envisager dans certains cas des recouvrements lorsque ceux-ci possèdent une structure plus riche permettant une écriture plus simple du modèle. Les concepts de décomposition (avec ou sans recouvrement) de systèmes et de classification des données ont été largement étudiés et appréhendés de diverses manières lesquelles ne sont pas exclusives mais plutôt complémentaires. La motivation de ce chapitre, n'est pas de présenter une étude exhaustive sur un sujet largement traité dans la littérature, mais tout simplement de rechercher une procédure de regroupement-simplification conduisant à une écriture aussi simple que possible des règles initiales.

L'extraction des règles composées peut s'effectuer à l'aide de deux approches différentes. La première approche utilise l'opération de regroupement des lignes initialement contrastées, la seconde fait appel à une procédure qui consiste dans un premier temps à effectuer le regroupement des lignes du tableau initial appartenant à la même classe de la partition, l'opération de contraste n'intervenant que dans un deuxième temps. Ceci, nous amène à étudier la procédure de regroupement dans les deux cas. En effet, le but visé par cette opération est la génération d'une règle composée à partir de chaque classe de la partition finale. Celle-ci peut être obtenue à partir de l'application de deux opérateurs que l'on peut définir comme suit :

. Opérateur de contraste C_σ donné par la relation (III.10, chapitre III) qui s'écrit :

$$C_\sigma : \begin{array}{ccc} D & \longrightarrow & D \\ P(Y/X = \alpha) & \longrightarrow & R(Y/X = \alpha) = C_\sigma [P(X/Y = \alpha)] \end{array} \quad (IV.1)$$

. Opérateur de regroupement RG : en notant toujours D l'ensemble des distributions conditionnelles de probabilité, celui-ci est défini, dans le cas de deux lignes, par :

$$RG : \begin{array}{ccc} D^2 & \longrightarrow & D \\ \{P(Y/X = \alpha_i), P(Y/X = \alpha_k)\} & \longrightarrow & P(Y/X = \alpha_i \vee \alpha_k) \end{array} \quad (IV.2)$$

'Nous montrerons que la composition de ces deux opérateurs permettra de définir deux nouveaux opérateurs [Cf. § III et IV]. Pour l'instant, nous allons évaluer la perte d'information provoquée par le regroupement des lignes initiales et voir comment évoluent la vérité et l'efficacité des règles composées en fonction des règles initiales.

II - REGROUPEMENT DES LIGNES DU TABLEAU INITIAL

Considérons deux lignes i et k du tableau initial et appliquons l'opérateur de regroupement RG donné par la relation (IV.2). L'opérateur de regroupement consiste à supprimer les deux lignes i et k et à les remplacer par une seule ligne ayant pour distribution :

$$P(Y/X = \alpha_i \vee \alpha_k) = \left\{ \frac{P_{ij} + P_{kj}}{P_{i.} + P_{k.}}, j = 1, \dots, m \right\}$$

Pour évaluer la perte d'information provoquée par cette opération, nous allons utiliser successivement la distance du χ^2 et un indice de distance basé sur les concepts entropiques (information de SHANNON).

II.1 Perte d'information provoquée par l'opération de regroupement des lignes initiales

II.1.1 Utilisation de la distance du χ^2

. Avant regroupement des lignes i et k , celle-ci s'écrit :

$$\frac{1}{L} \chi_0^2 = \sum_j \sum_{l \neq i, k} \frac{P_{lj}^2}{P_{l.} P_{.j}} + \sum_j \frac{P_{ij}^2}{P_{i.} P_{.j}} + \sum_j \frac{P_{kj}^2}{P_{k.} P_{.j}} - 1$$

Après regroupement, une telle distance devient :

$$\frac{1}{L} \chi_1^2 = \sum_j \sum_{l \neq i \neq k} \frac{P_{lj}^2}{P_l P_j} + \sum_j \frac{(P_{ij} + P_{kj})^2}{(P_i + P_k) P_j} - 1$$

La variation du χ^2 est donnée par :

$$\frac{1}{L} (\chi_0^2 - \chi_1^2) = \Delta \chi^2 = \frac{P_i P_k}{P_i + P_k} \sum_{j=1}^m \frac{1}{P_j} (P_{j/i} - P_{j/k})^2 \quad (\text{IV.3})$$

On peut alors vérifier très facilement que $\Delta \chi^2 = 0$ dans le cas de l'équivalence distributionnelle

$$P_{j/i} = P_{j/k}, \quad \forall j \in \{1, \dots, m\}$$

$\Delta \chi^2$ sera d'autant plus élevé que les deux distributions

$$P(Y/X = \alpha_i) \text{ et } P(Y/X = \alpha_k)$$

présentent des profils différents. La comparaison de $\Delta \chi^2$ calculé et $\Delta \chi^2$ tabulé permet de juger si le regroupement occasionne ou non une perte d'information significative.

II.1.2 Utilisation de l'information de SHANNON

. Avant le regroupement des lignes i et k, l'information entre Y et X s'écrit :

$$\begin{aligned} I_0(Y:X) &= H_0(Y) - H_0(Y/X) \\ &= \sum_j \sum_{l \neq i \neq k} P_{lj} \text{Log} \frac{P_{lj}}{P_l P_j} + \sum_j \left(P_{ij} \text{Log} \frac{P_{ij}}{P_i P_j} + P_{kj} \text{Log} \frac{P_{kj}}{P_k P_j} \right) \end{aligned}$$

. Après regroupement, cette quantité devient :

$$\begin{aligned} I_0(Y:X) &= H_0(Y) - H_0(Y/X) \\ &= \sum_j \sum_{l \neq i \neq k} P_{lj} \text{Log} \frac{P_{lj}}{P_l P_j} + \sum_j \left((P_{ij} + P_{kj}) \text{Log} \frac{P_{ij} + P_{kj}}{(P_i + P_k) P_j} \right) \end{aligned}$$

La perte d'information provoquée par une telle opération est donnée par :

$$\Delta I = I_0(Y:X) - I_1(Y:X) = H_1(Y/X) - H_0(Y/X) = \Delta H$$

puisque $H_0(Y) = H_1(Y)$ (la distribution $P(Y)$ reste inchangée) d'où :

$$\Delta I = \Delta H = \sum_{j=1}^m \left(P_{ij} \text{Log} \frac{P_{ij} (P_i + P_k)}{P_i (P_{ij} + P_{kj})} + P_{kj} \text{Log} \frac{P_{kj} (P_i + P_k)}{P_k (P_{ij} + P_{kj})} \right) \quad (\text{IV.4})$$

. On peut montrer facilement que cette quantité est positive ou nulle. En effet, l'utilisation de la relation (I.63) permet de conclure que $\Delta H \geq 0$. Ceci s'explique par le fait que l'information échangée entre deux variables diminue lorsque on effectue un regroupement de deux ou plusieurs modalités de l'une d'entre elles.

. On peut montrer également [SBAI 1983] que le minimum de ΔH est atteint lorsque $P(Y/X = \alpha_i)$ et $P(Y/X = \alpha_k)$ présentent des profils identiques (équivalence distributionnelle).

. Selon la valeur de ΔH calculée et en se fixant un seuil de perte d'information limite, il est alors possible d'accepter ou de rejeter l'opération de regroupement.

II.1.3 Comparaison de ΔI et $\Delta \chi^2$

Les lignes susceptibles d'être regroupées sont celles qui présentent des profils identiques ou voisins. A cet effet, on montre dans le cas particulier où $P(Y/X = \alpha_i)$; $P(Y/X = \alpha_k)$ et $P(Y)$ ont des distributions voisines que ΔH et $\Delta \chi^2$ sont identiques à une constante près [Cf annexe IV.1) :

$$\Delta \chi^2 \cong 2 \text{Log } 2 \cdot \Delta I \quad (\text{IV.5})$$

Cette approximation n'est possible que pour les faibles valeurs de ΔH et de $\Delta \chi^2$. Dans le cas général et comme le signale M. VOLLE [1976] "Dans la pratique, à l'usage, on constate que l'utilisation de l'une ou l'autre notion conduit à des résultats analogues : c'est ainsi que, lorsqu'on procède à une agrégation selon la variance, les arbres obtenus dans les deux cas sont le plus souvent identiques". Cependant, l'hypothèse ΔH ou $\Delta \chi^2$ faible ne s'applique pas forcément lors d'un regroupement (faible sur contraste mais fort sur initial par exemple).

Pour la suite des calculs, on utilisera ΔH puisqu'il fait intervenir uniquement les distributions des lignes regroupées alors que $\Delta \chi^2$ fait appel en plus de ces deux distributions à la distribution moyenne de l'ensemble des lignes. Le calcul de cette distribution moyenne risque d'être fastidieux dans le cas du regroupement des lignes contrastées puisque :

$$r_j = \sum_{i=1}^n r_{ij} = \sum_{i=1}^n P_{ij} \left(\frac{1}{\sum_1^i P_{j/i}} \right)$$

Que deviennent les coefficients de vérité et d'efficacité de la règle composée, obtenue par le regroupement des lignes du tableau initial ?

II.2 Coefficient de vérité de la règle composée

Nous avons défini au chapitre III une règle comme l'information extraite sous forme symbolique à partir d'une ligne (ou colonne) du tableau contrasté. Or, si on prend comme règle les données sous leur forme brute, c'est-à-dire :

$$X = \alpha \quad \text{ALORS} \quad Y = \beta_1 \vee \beta_2 \vee \dots \vee \beta_m$$

alors les coefficients de vérité V_i , V_k et $V_{i \vee k}$ des lignes i , k et $i \vee k$ s'écrivent :

$$V_i = V_k = V_{i \vee k} = 1 \quad (\text{IV.6})$$

En l'absence de contraste (les règles extraites du tableau initial sont sous leur forme brute), le coefficient de vérité d'une règle composée est identique aux coefficients de vérité des règles initiales et ceci quelles que soient les lignes regroupées.

II.3 Coefficient d'efficacité de la règle composée

Si on note e_i , e_k et $e_{i \vee k}$ les coefficients d'efficacité des lignes i , k et $i \vee k$, on peut alors écrire par définition (Cf § IV.4.4. chapitre III) :

$$e_i = 1 - \frac{H(Y/X=\alpha_i)}{\text{Log } m} \quad (\text{IV.7})$$

$$e_k = 1 - \frac{H(Y/X=\alpha_k)}{\text{Log } m} \quad (\text{IV.8})$$

$$e_{i \vee k} = 1 - \frac{H(Y/X=\alpha_i \vee \alpha_k)}{\text{Log } m} \quad (\text{IV.9})$$

. Dans le cas particulier de l'équivalence distributionnelle, c'est-à-dire : $P(Y/X = \alpha_i) = P(Y/X = \alpha_k) = P(Y/X = \alpha_i \vee \alpha_k)$: $e_{i \vee k} = e_i = e_k$: l'efficacité des règles initiales est alors conservée.

. Dans le cas général $e_{i \vee k}$ est une fonction décroissante de ΔI (perte d'information provoquée par le regroupement). En effet :

$$\Delta I = (P_i + P_k) H(Y/X = \alpha_i \vee \alpha_k) - P_i H(Y/X = \alpha_i) - P_k H(Y/X = \alpha_k) \geq 0$$

En divisant par $\text{Log } m$, tous les termes de cette égalité et en tenant compte des relations précédentes, on peut alors écrire :

$$e_{ivk} = \frac{P_i}{P_i + P_k} e_i + \frac{P_k}{P_i + P_k} e_k - \frac{\Delta I}{(P_i + P_k) \text{Log } m} \quad (\text{IV.10})$$

où

$$\Delta e_{ivk} = \frac{\Delta I}{(P_i + P_k) \text{Log } m} \quad (\text{IV.11})$$

apparaît comme une perte d'efficacité. Elle est proportionnelle à la perte d'information provoquée par le regroupement des lignes i et k . Par ailleurs, en notant :

$$p = \frac{P_i}{P_i + P_k} \quad \text{et} \quad q = \frac{P_k}{P_i + P_k}$$

$$\text{on a alors :} \quad e_{ivk} \leq p e_i + q e_k \quad \text{avec } p + q = 1$$

relation qui signifie que l'efficacité de la moyenne est inférieure ou égale à la moyenne des efficacités des règles initiales.

III - OPERATION DE CONTRASTE APRES AGREGATION DES LIGNES

III.1 La nouvelle distribution contrastée :

Après la procédure de regroupement des lignes i et k , on obtient la distribution $P(Y/X = \alpha_i \vee \alpha_k)$ qu'il convient de contraster.

Le choix du seuil de contraste "optimal" qu'on note $\delta_{ivk} = \delta$, s'effectue en fonction des critères que nous avons développé au chapitre III.

L'opérateur de contraste, de niveau δ , est alors défini comme suit :

$$C_\delta = \begin{cases} \frac{P_i + P_k}{\delta} & \text{pour } j \in \{1, \dots, \delta\} \\ \sum_1 (P_{ij} + P_{kj}) & \\ 0 & \text{pour } j \in \{\delta+1, \dots, m\} \end{cases} \quad (\text{IV.12})$$

La composition des opérateurs donnés par (IV.1) et (IV.2) permet de définir un nouvel opérateur à partir duquel la nouvelle distribution contrastée est obtenue. Pour alléger l'écriture, nous noterons respectivement P et R les distributions conditionnelles de probabilité initiale et contrastée d'une ligne du tableau, un tel opérateur est défini comme suit :

$$\begin{aligned}
 \cdot C_\delta \circ RG : \quad D^2 &\longrightarrow D \\
 (P_i, P_k) &\longrightarrow C_\delta [RG (P_i, P_k)] = C_\delta (P_{i \vee k}) = R_{\delta i \vee k} = R_\delta \quad (IV.13)
 \end{aligned}$$

La distribution contrastée R_δ s'écrit en notant $J = \{j, j \in (1, \dots, m), \text{ tels que } : \beta_j \in B_\delta (\alpha_i \vee \alpha_k) \}$ avec $|J| = \delta$:

$$R_\delta = \begin{cases} r_{j/i \vee k} = \frac{P_{ij} + P_{kj}}{\sum_{j \in J} P_{ij} + P_{kj}} & \text{pour } j \in J \\ 0 & \text{pour } j \notin J \end{cases} \quad (IV.14)$$

A l'issue de cette étape, les deux types de règles suivantes peuvent être extraites :

$$\cdot \text{type 1 : } X = \alpha_i \vee \alpha_k \text{ ALORS } Y = \beta_1 \vee \beta_2 \vee \dots \vee \beta_\delta \quad (IV.15)$$

$$\cdot \text{type 2 : } X = \alpha_i \vee \alpha_k \text{ ALORS } Y = (\beta_1, r_{1/i \vee k}) \vee \dots \vee (\beta_\delta, r_{\delta/i \vee k})$$

Calculons à présent les coefficients de vérité et d'efficacité de telles règles composées.

III.2 Coefficient de vérité de la règle composée

Le calcul d'un tel coefficient s'opère de la même façon que le cas d'une seule ligne. Par définition, celui-ci s'écrit :

$$V(\delta) = 1 - U(\delta) \quad (IV.16)$$

Remarquons que $V(\delta) \leq V_{i \vee k} = 1$, et que $V(\delta) = V_{i \vee k} = 1$ si et seulement si $U(\delta) = 0$

Rappelons toutefois que le choix du seuil δ se fait dans un contexte plus global à l'aide d'une analyse multi-critère définie dans le chapitre précédent.

III.3 Coefficient d'efficacité de la règle composée

Pour alléger l'écriture, on notera P et R_δ les distributions

$$P(Y/X = \alpha_i \vee \alpha_k) \text{ et } R_\delta(Y/X = \alpha_i \vee \alpha_k).$$

Le coefficient d'efficacité $e(\delta)$ est alors défini par :

$$\begin{aligned} e(\delta) &= 1 - \frac{H(R_\delta)}{\text{Log } m} \\ &= \left(1 - \frac{H(P)}{\text{Log } m}\right) + \frac{H(P) - H(R_\delta)}{\text{Log } m} \end{aligned} \quad (\text{IV.17})$$

or $H(P) - H(R_\delta) = \Delta H_\delta$ est la perte d'information provoquée par l'opération de contraste. $e(\delta)$ s'écrit alors :

$$\begin{aligned} e(\delta) &= e_{ivk} + \frac{\Delta H_\delta}{\text{Log } m} = e_{ivk} + \Delta e_\delta \\ &= \frac{P_i}{P_i + P_k} e_i + \frac{P_k}{P_i + P_k} e_k - \Delta e_{ivk} + \Delta e_\delta \end{aligned} \quad (\text{IV.18})$$

Ainsi, il apparaît que l'opérateur $C_\delta \circ \text{RG}$ conduit à :

- . une perte d'efficacité provoquée par l'opération de regroupement : Δe_{ivk}
- . un gain d'efficacité occasionné par l'opération de contraste : Δe_δ

Nous constatons que Δe_δ est proportionnel à la perte d'information due au contraste, il est d'autant plus grand que ΔH_δ est élevé. Un tel gain d'efficacité peut se faire au détriment de la vérité de la règle lorsque δ est petit et $U(\delta)$ non négligeable.

IV - REGROUPEMENT DES LIGNES DU TABLEAU CONTRASTE

IV.1 La nouvelle distribution contrastée

Après la procédure de contraste, appliquée aux lignes du tableau initial, on obtient des distributions contrastées que l'on se propose de regrouper. D'un point de vue formel, une telle opération s'obtient par la composition des opérateurs C_σ et RG définis par les relations (IV.1) et (IV.2). En effet, ces deux opérateurs permettent de définir un nouvel opérateur à partir duquel la règle composée peut être extraite. Si on note σ_i et σ_k les niveaux de contraste des lignes i et k, un tel opérateur est défini comme suit :

$$\begin{aligned}
\text{RG} \circ \text{C}_{\sigma_i * \sigma_k} : \quad D^2 &\longrightarrow D \\
(P_i, P_k) &\longrightarrow \text{RG} [\text{C}_{\sigma_i}(P_i), \text{C}_{\sigma_k}(P_k)] = \text{RG} (R_{\sigma_i}, R_{\sigma_k}) = R_{\sigma_i * \sigma_k} = R_{\sigma}
\end{aligned}
\tag{IV.19}$$

avec : $\sigma_i * \sigma_k = \sigma_{ik} = \sigma = \sigma_i + \sigma_k - \text{card} [B_{\sigma_i}(\alpha_i) \cap B_{\sigma_k}(\alpha_k)]$ où

$$B_{\sigma_i}(\alpha_i) = \{ \beta_j, j = 1, \dots, \sigma_i \text{ tel que : } r_{j/i} > 0 \text{ et } r_{j/i} = 0 \ \forall j \in (\sigma_i + 1, \dots, m) \}$$

σ est fixé par σ_i et σ_k les niveaux de contraste des lignes i et k . Leur regroupement ne modifie pas les conclusions des règles de niveaux σ_i et σ_k auxquelles elles donnent naissance. La règle composée ainsi obtenue n'est autre que la réunion de ces dernières. Par ailleurs, si on note :

$$I = \{ j, j \in (1, \dots, m) \text{ tel que : } \beta_j \in B_{\sigma}(\alpha_i \vee \alpha_k) \} \text{ avec } \text{card}(I) = \sigma$$

la nouvelle distribution contrastée R_{σ} obtenue à partir de R_{σ_i} et R_{σ_k} s'écrit alors :

$$R_{\sigma} = \text{RG} \circ \text{C}_{\sigma} = \left\{ \frac{r_{ij} + r_{kj}}{r_i + r_k}, \quad j \in I \right\} \tag{IV.20}$$

La nouvelle distribution contrastée R_{σ} est obtenue à partir du regroupement des distributions de niveaux σ_i et σ_k et il convient alors d'évaluer la perte d'information provoquée par une telle opération.

IV.2 Perte d'information occasionnée par le regroupement

La perte d'information provoquée par le regroupement des lignes i et k contrastées peut être également mesurée, comme dans le cas précédent, par l'utilisation de la distance du χ^2 ou bien par la variation informationnelle ΔI ou ΔH :

$$\Delta \chi_{\sigma}^2 = \frac{r_k \cdot r_i}{r_i + r_k} \sum_{j \in I} \frac{1}{r_j} (r_{j/i} - r_{j/k})^2 \tag{IV.21}$$

$$\Delta I_{\sigma} = \sum_{j \in I} \left(r_{ij} \text{Log} \frac{r_{ij}}{r_i} \frac{r_i + r_k}{r_{ij} + r_{kj}} + r_{kj} \text{Log} \frac{r_{kj}}{r_k} \frac{r_i + r_k}{r_{ij} + r_{kj}} \right) \tag{IV.22}$$

La perte d'information étant évaluée, il reste à déterminer les coefficients de vérité et d'efficacité de la règle de niveau σ obtenue par l'opérateur $\text{RG} \circ \text{C}_{\sigma}$ de façon à les comparer avec ceux des règles initiales de niveaux σ_i et σ_k .

IV.3 Coefficient de vérité de la règle composée

Il se calcule à partir de la distribution $P(Y/X = \alpha_i \vee \alpha_k)$ et du niveau de contraste σ . En effet, dans la mesure où les deux règles de niveaux σ_i et σ_k regroupées ont été jugées conformes à une certaine réalité, elles imposent à la règle composée le niveau σ , afin que celle-ci puisse conserver (reproduire, couvrir) la réalité modélisée par les règles initiales susceptibles d'être transformées par l'opération de regroupement. Autrement dit, l'écriture de la règle composée à partir des règles initiales, c'est-à-dire :

$$X = \alpha_i \vee \alpha_k \text{ ALORS } Y = B_{\sigma_i}(\alpha_i) \vee B_{\sigma_k}(\alpha_k)$$

signifie que Y doit contenir aussi bien les modalités $\beta_j \in B_{\sigma_i}(\alpha_i)$ que les modalités $\beta_j \in B_{\sigma_k}(\alpha_k)$ d'où la nécessité d'imposer à la nouvelle distribution ligne le niveau de contraste σ . Le coefficient de vérité s'écrit alors :

$$V(\sigma) = 1 - U(\sigma) \quad (\text{IV.23})$$

$$= 1 - \left(\sum_{j \notin I} \frac{P_{ij} + P_{kj}}{P_{i.} + P_{k.}} \right) \frac{\text{Log}(m - \sigma) - H(Q)}{\text{Log } m}$$

avec $Q = Q(Y/X = \alpha_i \vee \alpha_k)$ la distribution des résidus après regroupement qui s'écrit :

$$Q = \begin{cases} \frac{P_{ij} + P_{kj}}{\sum_{j \notin I} (P_{ij} + P_{kj})} , & \text{pour } j \notin I \\ 0 , & \text{pour } j \in I \end{cases}$$

Exprimons à présent le coefficient de vérité $V(\sigma)$ de la règle composée en fonction des coefficients de vérité $V_i(\sigma_i)$ et $V_k(\sigma_k)$ des règles initiales.

Proposition 1 :

$$\left[\begin{array}{l} \forall V_i(\sigma_i) \text{ et } V_k(\sigma_k) \quad \text{on a :} \\ V(\sigma) \geq \frac{P_{i.}}{P_{i.} + P_{k.}} V_i(\sigma_i) + \frac{P_{k.}}{P_{i.} + P_{k.}} V_k(\sigma_k) \end{array} \right. \quad (\text{IV.24})$$

Preuve : Notons :

$$\bullet p = \frac{P_i}{P_i + P_k} \quad \text{et} \quad q = \frac{P_k}{P_i + P_k}$$

$$\bullet A = \sum_{\sigma+1}^m P_{j/i} \leq \sum_{\sigma_i+1}^m P_{j/i}$$

$$\bullet B = \sum_{\sigma+1}^m P_{j/k} \leq \sum_{\sigma_k+1}^m P_{j/k}$$

$$\bullet C = \sum_{\sigma+1}^m \frac{P_{ij} + P_{kj}}{P_i + P_k}$$

$$\text{On a : } C = pA + qB$$

$$V(\sigma_i) = 1 - U_i(\sigma_i) \Rightarrow p V_i(\sigma_i) = p - p U_i(\sigma_i) \dots \quad (1)$$

$$V(\sigma_k) = 1 - U_k(\sigma_k) \Rightarrow q V_k(\sigma_k) = q - q U_k(\sigma_k) \dots \quad (2)$$

$$V(\sigma) = 1 - U(\sigma) \dots \dots \dots \quad (3)$$

$$(1), (2) \text{ et } (3) \Rightarrow V(\sigma) - p V_i(\sigma_i) - q V_k(\sigma_k) = [p U_i(\sigma_i) + q U_k(\sigma_k) - U(\sigma)]$$

Or, l'utilité de l'information perdue est fonction décroissante de σ , on a alors :

$$p U_i(\sigma_i) + q U_k(\sigma_k) \geq p U_i(\sigma) + q U_k(\sigma)$$

Montrons que : $[p U_i(\sigma) + q U_k(\sigma) - U(\sigma)] \geq 0$

En remplaçant les utilités par leurs valeurs respectives, cette quantité s'écrit :

$$[] = pA \frac{\text{Log}(m - \sigma) - H(Q_i)}{\text{Log } m} + qB \frac{\text{Log}(m - \sigma) - H(Q_k)}{\text{Log } m} - C \frac{\text{Log}(m - \sigma) - H(Q)}{\text{Log } m}$$

$$\Rightarrow [] \text{Log } m = CH(Q) - pA H(Q_i) - qB H(Q_k) = C \left[H(Q) - \frac{pA}{C} H(Q_i) - \frac{qB}{C} H(Q_k) \right]$$

$$= C \left[\sum_{\sigma+1}^m \left(\frac{p P_{j/i} + q P_{j/k}}{C} \right) \text{Log} \left(\frac{p P_{j/i} + q P_{j/k}}{C} \right) + \frac{pA}{C} \sum \frac{P_{j/i}}{A} \text{Log} \frac{P_{j/i}}{A} + \frac{qB}{C} \sum \frac{P_{j/k}}{B} \text{Log} \frac{P_{j/k}}{B} \right]$$

Notons $\frac{P_{j/i}}{A} = x_j$ et $\frac{P_{j/k}}{B} = y_j$

$$[] \text{Log } m = C \left[-\sum \left(\frac{p^A}{C} x_j + \frac{q^B}{C} y_j \right) \text{Log} \left(\frac{p^A}{C} x_j + \frac{q^B}{C} y_j \right) + \frac{p^A}{C} \sum x_j \text{Log } x_j + \frac{q^B}{C} \sum y_j \text{Log } y_j \right]$$

Posons $\frac{p^A}{C} = a$ et $\frac{q^B}{C} = b$

$$[] \text{Log } m = C \left[-\sum_{\sigma+1}^m (a x_j + b y_j) \text{Log} (a x_j + b y_j) + a \sum_{\sigma+1}^m x_j \text{Log } x_j + b \sum_{\sigma+1}^m y_j \text{Log } y_j \right]$$

Or d'après la propriété de convexité de la fonction $-x \log x$, la quantité entre crochets est ≥ 0 . Par conséquent :

$$V(\sigma) \geq p V_i(\sigma_i) + q V_k(\sigma_k) \quad \text{C.Q.F.D.}$$

Interprétation :

Cette proposition montre que la fonction $V(\sigma)$ vérifie une propriété de concavité. Elle exprime le fait que le coefficient de vérité de la règle composée est supérieur ou égal à la moyenne pondérée des coefficients de vérité des règles initiales. Ceci s'explique par le fait que $U(\sigma)$ calculé à partir de la distribution moyenne des résidus $Q (Y/X = \alpha_i \vee \alpha_k)$ est inférieur à $p U_i(\sigma) + q U_k(\sigma)$ calculé à partir des distributions $Q (Y/X = \alpha_i)$ et $Q (Y/X = \alpha_k)$. Par conséquent, le regroupement des règles dans ce cas ne s'accompagne d'aucune perte de vérité mais bien au contraire un gain de vérité peut être obtenu lorsque $Q (Y/X = \alpha_i)$ et $Q (Y/X = \alpha_k)$ sont quelconques mais $Q (Y/X = \alpha_i \vee \alpha_k)$ a tendance à se répartir de façon équiprobable. Dans le cas limite où cette distribution est équiprobable, on a : $V(\sigma) = 1$.

En définitive, nous constatons à travers cette proposition que le regroupement des règles n'occasionne aucunement la dégradation du coefficient du vérité de la règle composée ainsi déduite. C'est pourquoi, nous laisserons volontairement de côté, le coefficient de vérité, pour éviter d'augmenter inutilement le temps de calcul de la procédure de simplification des règles.

IV.4 Coefficient d'efficacité de la règle composée

Si on note P et R_σ les distributions $P(Y/X=\alpha_i \vee \alpha_k)$ et $R_\sigma (Y/X=\alpha_i \vee \alpha_k)$, le coefficient d'efficacité $e(\sigma)$ est alors défini par :

$$\begin{aligned}
e(\sigma) &= 1 - \frac{H(R_\sigma)}{\text{Log } m} & (IV.25) \\
&= \left(1 - \frac{H(P)}{\text{Log } m}\right) - \frac{H(P) - H(R_\sigma)}{\text{Log } m} = e_{i \vee k} - \frac{\Delta H_\sigma}{\text{Log } m}
\end{aligned}$$

Comme dans le cas précédent, $H(P) - H(R_\sigma) = \Delta H_\sigma$ représente la quantité d'information perdue globalement par les deux opérations de contraste sur les lignes i et k . $e(\sigma)$ peut s'exprimer également en fonction de $e(\sigma_i)$ et $e(\sigma_k)$ les coefficients d'efficacité des règles initiales. En effet, si on note ΔI_σ la quantité d'information perdue par le regroupement des lignes i et k contrastées, $e(\sigma)$ peut s'écrire également :

$$e(\sigma) = \frac{P_i}{P_i + P_k} e(\sigma_i) + \frac{P_k}{P_i + P_k} e(\sigma_k) - \frac{\Delta I_\sigma}{(P_i + P_k) \text{Log } m} \quad (IV.26)$$

où

$$\frac{\Delta I_\sigma}{(P_i + P_k) \text{Log } m} = \Delta e$$

apparaît comme une perte d'efficacité. Elle est proportionnelle à la perte d'information provoquée par le regroupement des lignes i et k contrastées. Par ailleurs, si on tient compte des pertes d'informations $\Delta H_i = H(P_i) - H(R_{\sigma_i})$ et $\Delta H_k = H(P_k) - H(R_{\sigma_k})$ causées respectivement par les opérations de contraste sur les lignes i et k , $e(\sigma)$ s'écrit alors :

$$\begin{aligned}
e(\sigma) &= \frac{P_i}{P_i + P_k} e_i + \frac{P_k}{P_i + P_k} e_k - \Delta e + \Delta e_\sigma \\
&\text{où } e_i \text{ et } e_k \text{ sont les efficacités des lignes initiales et} & (IV.27) \\
\Delta e_\sigma &= \frac{P_i \Delta H_i + P_k \Delta H_k}{P_i + P_k}
\end{aligned}$$

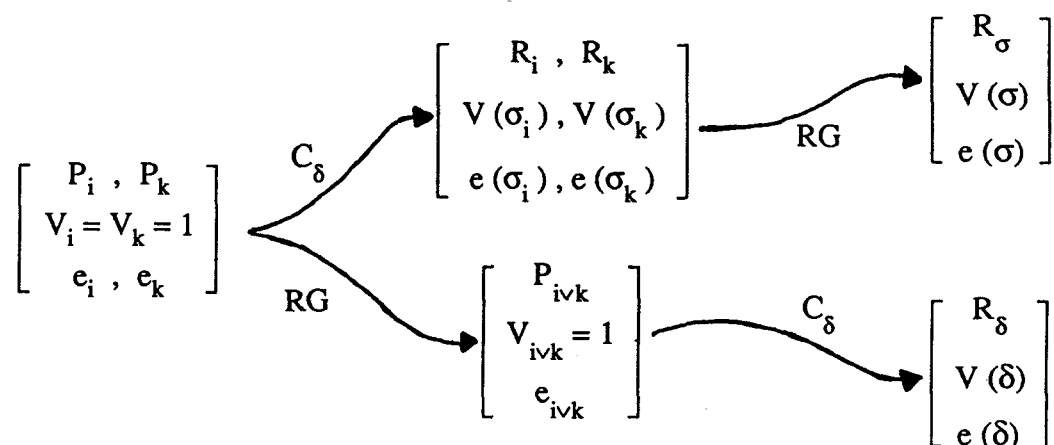
apparaît comme un gain d'efficacité. Il est proportionnel à la quantité d'information perdue par les opérations de contraste.

Le calcul de $e(\sigma)$ montre bien que l'opération de contraste s'accompagne d'un gain d'efficacité alors que l'opération de regroupement provoque une perte de celle-ci.

En définitive, nous avons montré qu'à partir de deux lignes du tableau initial, il est possible de générer une règle composée soit par l'utilisation de l'opérateur RG o C_σ ou bien par l'opérateur C_δ o RG . Il serait donc utile de comparer les deux procédures pour voir dans quelle mesure celles-ci peuvent ou non donner naissance à la même règle.

V - COMPARAISON DES DEUX PROCEDURES

Les deux procédures d'extraction des règles composées peuvent être résumées comme suit :



A l'issue de cette étape, les deux règles composées susceptibles d'être extraites s'écrivent comme suit :

$$. X = \alpha_i \vee \alpha_k \text{ ALORS } Y \in B_\sigma(\alpha_i \vee \alpha_k)$$

$$. X = \alpha_i \vee \alpha_k \text{ ALORS } Y \in B_\delta(\alpha_i \vee \alpha_k)$$

La comparaison de ces deux règles nécessite la comparaison des grandeurs suivantes :

. les coefficients d'efficacité $e(\sigma)$ et $e(\delta)$: ceci revient en réalité à comparer les gains et les pertes d'efficacité dans les deux cas, qui se traduisent par des pertes d'information causées conjointement par l'opération de contraste et de regroupement.

. les coefficients de vérité $V(\sigma)$ et $V(\delta)$: ceux-ci dépendent essentiellement des utilités de l'information perdue aux niveaux σ et δ .

La comparaison de ces grandeurs s'opère à partir des distributions R_σ et R_δ obtenues par les opérateurs $RG \circ C_\sigma$ et $C_\delta \circ RG$.

V.1 Comparaison de $RG \circ C_\sigma$ et $C_\delta \circ RG$

En se référant aux indices I et J définis aux § III.1 et IV.1, on peut alors énoncer la proposition suivante :

Proposition 2 :

$$\left[\begin{array}{l} \forall P(Y/X=\alpha_i) \text{ et } P(Y/X=\alpha_k) \\ \text{si } B_{\sigma}(\alpha_i \vee \alpha_k) = B_{\delta}(\alpha_i \vee \alpha_k) \text{ et } \sum_{j \in I} P_{j/i} = \sum_{j \in J} P_{j/k} \\ \text{alors } RG \circ C_{\sigma} = C_{\delta} \circ RG \end{array} \right.$$

Preuve :

$$\bullet RG \circ C_{\sigma} = R_{\sigma} = \left\{ \frac{r_{ij} + r_{kj}}{r_{i.} + r_{k.}}, j \in I \right\}$$

$$\frac{r_{ij} + r_{kj}}{r_{i.} + r_{k.}} = \frac{1}{P_{i.} + P_{k.}} \frac{1}{\sum_{j \in I} P_{j/i}} (P_{ij} + P_{kj})$$

$$\bullet C_{\delta} \circ RG = \left\{ \frac{P_{ij} + P_{kj}}{\sum_{j \in J} (P_{ij} + P_{kj})}, j \in J = I \right\}$$

$$\frac{P_{ij} + P_{kj}}{\sum_j (P_{ij} + P_{kj})} = \frac{P_{ij} + P_{kj}}{P_{i.} \sum_j P_{j/i} + P_{k.} \sum_j P_{j/k}} = \frac{1}{P_{i.} + P_{k.}} \frac{1}{\sum_{j \in I} P_{j/i}} (P_{ij} + P_{kj})$$

d'où $C_{\delta} \circ RG = RG \circ C_{\sigma}$

Interprétation :

Si les résidus contenus dans l'intervalle $\{j, j \in I = J\}$ ne contiennent aucune information opérationnelle, les deux opérateurs précédents conduisent à une seule et unique règle composée.

Dans la mesure où les niveaux de contraste σ_i, σ_k et δ sont choisis en fonction de critères multiples, il est difficile d'établir une comparaison entre eux dans le cas général d'autant plus que les distributions R_{σ} et R_{δ} auxquelles ils aboutissent ne sont pas identiques. Ceci rend également difficile la comparaison de $e(\sigma)$ et $e(\delta)$ qui dépendent principalement des distributions R_{σ} et R_{δ} . Il en est de même pour les coefficients de vérité puisque ceux-ci se déterminent à partir de la distribution $P(Y/X = \alpha_i \vee \alpha_k)$ et des niveaux σ et δ . On peut cependant faire les remarques suivantes :

- si $\delta \leq \sigma$ alors $V(\sigma) \geq V(\delta)$: en effet, $V(\sigma) = 1 - U(\sigma) \geq 1 - U(\delta)$ puisque U est fonction décroissante du niveau de contraste (cf. chapitre III § IV.4.3.5.)
- d'un point de vue pratique, même si $\delta \leq \sigma$, $V(\sigma)$ et $V(\delta)$ seront approximativement identiques à cause de l'information perdue, dans l'intervalle $I \setminus J$, qui a été jugée inutile par l'opérateur C_δ o RG. Par contre, dans ce cas, la règle de niveau δ sera plus efficace que celle de niveau σ .

VI - SIMPLIFICATION DES REGLES COMPOSEES

VI.1 Introduction

Dans les paragraphes précédents, nous avons montré qu'une voie possible permettant la diminution du nombre de règles est de procéder à leur regroupement. Une telle opération peut se traduire également par une diminution de la complexité de ces règles. En effet, l'évaluation de la partie hypothèse des règles sera d'autant plus facilement obtenue que celle-ci se présente sous une forme plus simple.

Soit $P_K(X) = \{A_1, A_2, \dots, A_K\}$ une partition de M_X en K classes, avec $\text{card}(A_i) = k_i$. Chaque classe $A_i \in P_K(X)$ peut s'écrire comme suit :

$$A_i = \bigvee_{j=1}^{k_i} \bigwedge_{l=1}^p \alpha_j^l \quad (\text{IV.29})$$

En effet, on rappelle que X est un vecteur à p composantes, c'est-à-dire : $X = \{X_i, i = 1, \dots, p\}$, α_j^p désigne alors la modalité prise par la composante X_1 sur la $j^{\text{ème}}$ ligne du sous tableau A_i .

A chaque partie hypothèse $A_i, i = 1, \dots, k$ de la règle, on fait correspondre la partie conclusion $B_{\sigma_i}(A_i) = B_i$ avec $\text{card}(B_i) = \sigma_i$.

L'ensemble des règles de la partition $P_K(X)$ s'écrivent :

$$X = A_i \text{ ALORS } Y = B_i \quad \forall i \in \{1, \dots, K\} \quad (\text{IV.30})$$

$X = A_i$ se présente comme une fonction combinatoire multivaluée qu'il est éventuellement possible de simplifier.

VI.2 Les techniques habituelles de simplification

VI.2.1 Les techniques graphiques [méthode des tableaux de KARNAUGH]

C'est la méthode la plus connue, elle consiste à représenter graphiquement dans une table de vérité les termes produits décrivant la fonction à simplifier. La recherche des termes adjacents conduit à la simplification d'une telle fonction. Or cette méthode est à écarter puisqu'elle est difficilement utilisable lorsque le nombre de variables et de modalités devient élevé et de plus elle est non programmable.

VI.2.2 Les techniques algébriques ou systématiques [méthode de QUINE MC CLUSKEY, principe PRESTO]

Cette méthode peut être également utilisée, elle consiste à décomposer les termes de la fonction en classes et à rechercher ensuite les adjacences classe par classe jusqu'à ce que celles-ci deviennent stables.

Cette méthode présente l'avantage d'être programmable mais malheureusement le temps de calcul s'accroît exponentiellement avec le nombre de variables. Pour réduire ce temps de calcul, on utilise parfois dans le cas binaire une méthode d'approximation appelée : principe PRESTO. A cet effet, PRESTO crée deux fonctions F_0 et F_1 décrites dans l'algèbre de BOOLE qui représentent sous forme de tables les équations d'origine. Au départ F_0 et F_1 sont identiques, on élimine temporairement chaque terme produit dans la table F_1 et on vérifie que la fonction résultante est toujours couverte par F_0 . Dans l'affirmative, le terme produit sera éliminé de F_1 . Le processus de réduction continue jusqu'à ce que F_1 soit stable [PRESTO 1985]. La solution obtenue contient forcément la fonction réduite, mais elle n'est pas nécessairement optimale, puisqu'elle peut contenir un ou plusieurs termes de plus que la solution minimale. Ce principe, tout en générant des solutions sous-optimales réduit considérablement le temps global de traitement. Le problème de simplification des règles peut être traité de façon similaire à celui des fonctions combinatoires. La seule différence qui existe dans les deux formalismes réside dans le fait que dans le deuxième cas, une variable est définie dans l'algèbre de BOOLE (variables à 2 modalités) alors que dans le premier cas, une variable est définie dans l'algèbre de POST (variables à modalités multiples). En 1921, POST a introduit une algèbre permettant l'étude des propositions logiques multivaluées. L'étude et les propriétés d'une telle algèbre sont données en annexe IV.2. A partir de ces travaux, différents auteurs ont proposé diverses représentations des algèbres de POST isomorphes les unes des autres, les plus connues sont :

- la représentation disjointe de l'algèbre de POST introduite par TRACZYCK et BRADDOCK en 1971 [TRACZYCK 71] ;

- le système monotone introduit également par TRACZYCK et utilisé par BRADDOCK et EPSTEIN pour définir une nouvelle algèbre appelée parfois : algèbre de MUEHLORLF qui est un treillis distributif avec un élément nul, un élément universel et vérifiant certains axiomes. On démontre que ce système est isomorphe à une algèbre de POST [NUTTER 74] ;
- le système algébrique libre dû à BRADDOCK et EPSTEIN [1971] ;

Cette représentation libre est également un treillis distributif avec un élément neutre et un élément unité et vérifiant certains axiomes qui font que cette algèbre est isomorphe à une algèbre de POST [NUTTER 74].

- le système de HERMANN généralisé : système introduit par HERMANN pour une algèbre binaire et généralisé ensuite dans le cas multi-valué [HERMANN 70] ;
- les φ -algèbres : c'est une généralisation des algèbres de POST qui a été présentée par BOSSUET en 1977. Celui-ci montre que parmi les $m!$ φ -algèbres correspondant à un ensemble m -valué, il est possible de trouver une φ -algèbre conduisant à une implémentation minimale avec un ensemble d'opérateurs unaires donnés. Dans ce sens, D. ETIEMBLE [1979] utilise les φ -algèbres pour obtenir une synthèse minimale des fonctions m -valuées.

Par ailleurs, J. DUSSAULT [1976] montre qu'il est possible, à partir d'un codage booléen d'une variable de POST, de définir une algèbre multi-valuée avec des propriétés booléennes. La seule différence entre les deux procédures provient des structures des treillis sur lesquelles les variables sont définies. En effet, le treillis booléen est distributif et complété alors que le treillis de POST est seulement distributif.

En définitive, l'ensemble des règles susceptibles d'être simplifiées se présentent comme des fonctions de POST multi-valuées ; la synthèse optimale de telles fonctions est alors possible [BOSSUET 77] [ETIEMBLE 79]. Une telle simplification conduit, pour la partition de M_X considérée, à des fonctions minimales multi-valuées. Nous allons montrer dans ce qui suit que, dans certains cas, l'utilisation de l'entropie du Max peut conduire également à la simplification de telles fonctions et permet, par conséquent, l'écriture des règles sous une forme réduite, fournissant ainsi un modèle simplifié du système.

VI.3 Méthode entropique de simplification

C'est une méthode basée essentiellement sur l'utilisation de l'entropie du Max. Mais avant d'exposer les principes de base d'une telle méthode de simplification, nous commencerons par un exemple introductif.

VI.3.1 Exemple

Soit le tableau de données ci-dessous dans lequel Y, X₁, X₂ et X₃ sont décrites dans l'algèbre de BOOLE. Pour rechercher une représentation simplifiée d'un tel système, utilisons respectivement l'entropie de SHANNON et l'entropie du Max.

X ₁	X ₂	X ₃	Y
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

Tableau IV.1

	Max	Shannon
H (Y/X ₁) =	0	0,4056
H (Y/X ₂) =	1/8	0,9056
H (Y/X ₃) =	1/8	0,9056
H (Y/X ₁ X ₂) =	0	0,25
H (Y/X ₁ X ₃) =	0	0,25
H (Y/X ₂ X ₃) =	0	0,25
H (Y/X ₁ X ₂ X ₃) =	0	0

Tableau IV.2

Nous constatons que l'entropie de SHANNON ne permet de conclure que Y est totalement expliqué que lorsque l'ensemble des variables sont prises en compte. Les seules règles qu'elle peut exprimer sont les règles atomiques contenues dans le tableau de données, c'est-à-dire :

$$X_1 X_2 X_3 = 000 \text{ alors } Y = 0, \dots, X_1 X_2 X_3 = 111 \text{ alors } Y = 1$$

Or, l'utilisation de l'entropie du Max permet de retrouver l'ensemble des règles, aussi bien les règles atomiques que les règles écrites sous une forme simplifiée. Ainsi, si on se réfère au tableau précédent :

$$H(Y/X_1) = 0 \text{ signifie : } X_1 = 1 \text{ alors } Y = 1$$

$$H(Y/X_2 X_3) = 0 \text{ signifie : } X_2 X_3 = 11 \text{ alors } Y = 1$$

$$H(Y/X_1 X_2) = 0 \text{ signifie : } X_1 X_2 = 00 \text{ alors } Y = 0$$

$$H(Y/X_1 X_3) = 0 \text{ signifie : } X_1 X_3 = 00 \text{ alors } Y = 0$$

$$H(Y/X_1 X_2 X_3) = 0 \text{ signifie : } X_1 X_2 X_3 = 0 \text{ alors } Y = 0 \dots \text{ (les règles atomiques)}$$

En réalité, le tableau de données analysé correspond à la fonction logique combinatoire suivante :

$$\left\{ \begin{array}{l} Y = X_1 + X_2 X_3 \\ \text{ou bien} \\ \bar{Y} = \bar{X}_1 \bar{X}_2 + \bar{X}_1 X_3 \end{array} \right.$$

Autrement dit, l'analyse des résultats de l'entropie du Max conduit à l'identification de l'ensemble des monômes générateurs de la fonction logique donc des règles simplifiées auxquelles ils donnent naissance.

Par ailleurs, l'entropie de SHANNON peut être utilisée conjointement avec l'entropie du Max dans la procédure de recherche des monômes générateurs de la fonction logique. En effet, si on note S_K un sous-ensemble de variables, avec $\text{card}(S^K) = K$, le calcul de $H_S(Y/S^K)$ permet de savoir si S^K contient ou non l'ensemble des monômes générateurs. Autrement dit, si $H_S(Y/S^K) = 0$, cela signifie que S^K contient l'ensemble des variables explicatives de Y et il est inutile de chercher au moyen de l'entropie du Max des monômes générateurs non issus de S^K . Ainsi, l'entropie de SHANNON peut être utilisée conjointement avec l'entropie du Max pour tester si toutes les variables explicatives ont été prises en compte (test d'arrêt).

VI.3.2 Méthode de simplification

Nous allons proposer deux théorèmes qui vont préciser dans quelles conditions les règles du type $X = \alpha$ ALORS $Y = \beta$ peuvent être extraites d'un tableau contenant l'ensemble des règles du système écrites sous leur forme redondante (règles atomiques).

a) Théorème 1

$$\begin{aligned} &\forall S \in \mathcal{P}(X) \\ &\text{si } H_M(Y/S) = 0 \\ &\text{alors } \exists \text{ au moins } (\alpha_i, \beta_j) \text{ tel que : } P(Y = \beta_j / S = \alpha_i) = 1 \end{aligned} \quad (\text{IV.31})$$

Preuve :

$$\text{Notons } P_{II} = \max_{i,j} P_{ij} \quad , \quad P_I = \max_i P_i.$$

$$\text{et } (IJ)^* = \{ (i,j) \text{ tel que : } p_{ij} = P_{II} \}$$

$$\text{Soit } (i^*, j^*) \in (IJ)^*$$

$$P_{i^* j^*} = \max_{i,j} P_{ij} = P_{II} = P_I \quad (\text{par hypothèse})$$

Par ailleurs :

$$P_{i^*} = P_{i^* j^*} + \sum_{j \neq j^*} P_{i^* j} \geq P_{i^* j^*} = P_I$$

$$\text{d'où } P_{i^*} = P_I \text{ et } P_{i^* j} = 0 \quad \forall j \neq j^*$$

Donc $P_{i^*} = P_{i^*j^*} \Rightarrow P(Y = \beta_{j^*} / S = \alpha_{i^*}) = 1$ C.Q.F.D.

Interprétation :

$H_M(Y/S) = 0$ signifie qu'il existe au moins une règle écrite sous forme simplifiée puisque S est de cardinal plus faible que X . Mais la réciproque d'un tel théorème n'est pas toujours vraie.

b) Théorème 2 :

$\forall S \in \mathcal{P}(X)$

si $\exists i^*$ tel que : $P(Y = \beta_j / S = \alpha_{i^*}) = 1$ avec $P_{i^*} = \max_i P_i$

alors $H(Y/S) = 0$ (IV.32)

Preuve :

$\exists i^*$ tel que $P(Y = \beta_j / S = \alpha_{i^*}) = 1 \Rightarrow \exists j^*$ tel que $P_{i^*j^*} = P_{i^*}$

$H_M(S) = 1 - \max_i P_i = 1 - P_{i^*}$ par hypothèse

$H_M(S \cup Y) = 1 - \max_{i,j} P_{ij} = 1 - P_{i^*j^*}$, en effet

$\forall i \neq i^* \quad P_i = \sum_j P_{ij} \leq P_{i^*}$

$\forall i, j, \quad i \neq i^* : P_{ij} \leq P_{i^*} = P_{i^*j^*}$

Pour $i = i^*$ et $j \neq j^* : P_{i^*j} = 0$ car $P_{i^*j^*} = P_{i^*}$

Donc $\forall i, j : P_{ij} \leq P_{i^*j^*}$ d'où :

$H_M(Y \cup S) = 1 - P_{i^*j^*}$ et $H_M(Y/S) = 0$ C.Q.F.D.

Cas particulier :

Si S possède n modalités équiprobables :

$$P_i = \frac{1}{n} = \max_k P_k, \quad \forall i$$

et le théorème précédent devient :

$$\begin{aligned} & \forall S \in \mathcal{P}(Y) \\ & \text{si } \exists i^* \text{ tel que } P(Y = \beta_j / S = \alpha_{i^*}) = 1 \\ & \text{alors } H_M(Y/S) = 0 \end{aligned} \tag{IV.33}$$

La preuve est identique à celle du théorème 2.

Interprétation

$\forall S \in \mathcal{P}(X)$ tel que $\exists (i,j)$ vérifiant $P(Y = \beta_j / S = \alpha_i) = 1$, on aura $H_M(Y/S) = 0$. En particulier, considérons une règle composée telle que :

$$(S_1 = \alpha_1) \vee (S_2 = \alpha_2) \vee \dots \vee (S_k = \alpha_k) \Rightarrow Y = \beta$$

Sous la condition d'équiprobabilité ci-dessus, on trouvera :

$$H_M(Y/S_i) = 0 : \quad \forall i \in (1, \dots, k)$$

alors que $H_S(Y/S_i)$ est en général non nul.

Un tel résultat montre que dans le cas où les modalités des variables explicatives sont équiprobables, toutes les règles qui constituent le modèle du système sont déduites. Un tel modèle contient aussi bien les règles initiales du tableau de données (les règles atomiques) que les règles écrites sous une forme simplifiée. Si une simplification est possible, cela veut dire qu'il existe des règles redondantes qu'il convient de simplifier.

c) Identification des règles redondantes :

L'utilisation de l'entropie du Max conduit à l'identification de tous les monômes de la règle. Il s'agit alors d'identifier ceux qui sont redondants, ce qui permettra d'écrire la règle sous sa forme la plus simple. En effet, les règles composées s'écrivent initialement (cf relation IV.25) :

$X = A_i$ ALORS $Y = B_i$ où $X = A_i$ est la fonction multivaluée à simplifier, qui est composée initialement d'un ensemble de monômes redondants. L'entropie du Max, détermine aussi bien ces monômes que les monômes simplifiés. Si on note \mathcal{M} l'ensemble de tous les mo-

monômes simplifiés ou non, celui-ci s'écrit :

$$S_i = \{S, S \in \mathcal{P}(X) \text{ tel que : } \exists \alpha \in M_S \text{ pour lequel } P(Y = B_i / S = \alpha) = 1\}$$

Parmi les éléments de cet ensemble, ceux qui fournissent les monômes minimaux (non redondants) sont définis par :

$$S_i^* = \{S, S \in S_i \text{ et tel que : } \forall S' \subset S \text{ et } \forall \alpha' \in M_{S'}, P(Y = B_i / S' = \alpha') < 1\}$$

Les monômes redondants à éliminer, s'écrivent sous la forme $S \cup R$ avec $S \in S_i^*$ et $R \in \mathcal{P}(X \setminus S^*)$.

VI.3.3 Exemple d'application

Considérons un système de quatre variables décrit dans la logique ternaire et donné par le tableau IV.3 de données ci-dessous, dans lequel Y et $X = \{X_1, X_2, X_3\}$ prennent les modalités 0, 1 et 2.

Ces données peuvent être schématisées par le tableau de KARNAUGH suivant (tableau IV.4) :

Les monômes s'écrivent : $S_1 = X_1$

$$S_2 = X_1 X_2, \quad S_3 = X_1 X_3$$

$$S_4 = X_1 X_2 X_3$$

Après simplification, on peut écrire les règles suivantes (cf tableau IV.6 ci-dessous)

$$X = X_1^0 + X_1^1 X_3^2 \Rightarrow Y = 0$$

$$X = X_1^1 X_3^0 + X_1^1 X_3^1 \Rightarrow Y = 1$$

$$X = X_1^2 \Rightarrow Y = 2$$

VL.4 Algorithme de simplification du modèle

Nous avons montré dans les paragraphes précédents qu'il était possible de regrouper les règles atomiques. Celles-ci conduisent à une partition de l'ensemble des modalités de X et s'écrivent :

$$X = A_i = \bigvee_{i=1}^l \alpha_i^1 \quad \text{ALORS} \quad Y = B_i$$

Nous avons montré également qu'une telle opération peut s'accompagner d'une simplification de ces règles. En effet, dans ce dernier cas :

$$X = A_i = \bigvee_{i=1}^l \alpha_i^1 \quad \text{ALORS} \quad Y = B_i$$

fait intervenir $a_i = k_i \cdot p$ modalités. Après application des méthodes de simplification, une telle expression ne comptera plus que $\bar{a}_i \leq a_i$ modalités. Pour mesurer la qualité de cette opération, on définit le coefficient suivant :

$$s(A_i) = \frac{a_i - \bar{a}_i}{a_i - 1} \quad (\text{IV.34})$$

que l'on appellera : coefficient simplificateur du regroupement A_i .

- si $\bar{a}_i = a_i$ alors $s(A_i) = 0$: aucune simplification n'est opérée
- si $\bar{a}_i = 1$ alors $s(A_i) = 1$: A_i est simplifié au maximum et se réduit à une modalité d'une variable $X_i \in X$; $i \in \{1, \dots, p\}$.

Nous remarquons que si $a_i = 1$, la partie hypothèse étant exprimée sous sa forme la plus simple possible, le problème de simplification ne se pose évidemment pas. Une telle simplification aura pour conséquence une diminution notable du nombre des variables d'entrée à mesurer. Cette façon de procéder conduit ainsi à diminuer le coût du test dans la phase utilisation du modèle. En effet, la simplification conduit à retenir pour l'explication d'une ou plusieurs modalités de Y , une partition de M_X plus grosse que la partition atomique. Soit X' une variable explicative associée à une partition $P(M_{X'}) > P(M_X)$ (où le signe $>$ signifie plus grosse que (cf § II.2 chapitre I) obtenu par regroupement et simplification d'un certain nombre de règles. Deux situations peuvent se rencontrer :

i) $P(M_{X'}) = P(M_S)$ avec $S \in \mathcal{P}(X)$ et $S \neq X$

Cette situation résulte soit d'une indépendance de la conclusion par rapport aux variables de l'ensemble $(X \setminus S)$, soit d'une redondance des variables explicatives ; l'information ap-

portée par $(X \setminus S)$ étant déjà fournie par S lorsque $Y = B_i$. En effet :

$$Y \in B_i \Rightarrow H(Y/S) = H(Y/X)$$

$$\text{ii) } P(M_X) \neq P(M_S), \forall S \in \mathcal{P}(X)$$

$P(M_X) > P(M_S)$ correspond au fait que pour une ou plusieurs des modalités de X , une partition plus grosse de leur ensemble de modalités donne la même information au sujet de Y que leur partition atomique. Ceci correspond à une indépendance locale de Y par rapport à ces variables.

Dans les deux cas, la simplification est obtenue soit par l'application des méthodes issues de l'algèbre de POST, soit par l'utilisation de la méthode entropique. Par ailleurs, nous avons montré que l'opération de regroupement des lignes (initiales ou contrastées) occasionne une perte d'information qui se traduit par une perte d'efficacité des règles initiales. On peut remarquer cependant, comme nous l'avons signalé à la proposition 1, que l'opération de regroupement conduit à une règle composée dont le coefficient de vérité ne se trouve aucunement dégradé. Dans le cas le plus défavorable :

$$V(\sigma) = \sum_1^{\sigma} P_{j/i} \quad \text{qui est le coefficient de vérité probabiliste.}$$

$$\text{Or } \sum_1^{\sigma} P_{j/i}$$

est l'un des critères retenus pour le choix du niveau de contraste, qui doit être aussi élevé que possible. C'est pourquoi, nous nous limiterons, pour juger les règles à regrouper, au coefficient d'efficacité.

En définitive, à chaque recouvrement de M_X , on peut associer un ensemble de règles qui constituent un modèle du système. Le modèle simplifié du système est obtenu en cherchant des règles composées dont l'efficacité n'est pas trop dégradée par rapport à l'efficacité des règles atomiques qu'elles remplacent, mais dont la partie hypothèse s'exprimera sous une forme plus simple. Il s'agit alors de chercher un recouvrement de l'ensemble des lignes du tableau par des classes A_i , la règle simplifiée correspondant à chaque classe est évaluée par :

- la perte d'efficacité qu'occasionne le regroupement ;
- le gain de simplicité qu'il permet.

Si on note ϵ_R un seuil de perte d'efficacité limite toléré par le regroupement et ϵ_S un seuil de gain de simplicité, l'ensemble des règles globales retenues est défini par les classes A_i telles que :

$$\Delta e (A_i) \leq \epsilon_R$$

$$s (A_i) \geq \epsilon_S$$

Les règles atomiques telles que α_i n'appartient à aucun des A_i retenus, restent sous leur forme initiale.

Ces classes A_i et donc le recouvrement qui en résulte sont obtenues par la mise en oeuvre de l'algorithme suivant :

Algorithme :

Soit V_i l'ensemble des voisins de la modalité α_i au sens de la variation de l'efficacité.

$$W_1 = \{\alpha_i, i = 1, \dots, n\}$$

$i = 1$

tant que $W_i \neq \emptyset$

faire $V_i = \{\alpha_j, j \neq i, j \in (1, \dots, n) \text{ tel } \Delta e (\sigma_{ij}) \leq \epsilon_R\}$

$C_i = \alpha_i \cup \{ \alpha, \text{ tel que } \forall (\alpha_j, \alpha_k) \in V_i : \Delta e (\sigma_{jk}) \leq \epsilon_R \}$: classe d'équivalence formée par α_i et un sous-ensemble de ses voisins

$$W_i = \left\{ \alpha_i, \text{ tel que } \alpha_i \in \bigcup_{j=1}^i C_j \right\}$$

l'ensemble des α_i n'appartenant à aucune classe C_i .

fin faire

fin tant que

Parmi l'ensemble des recouvrements possibles des classes A_i , on choisit celui qui présente le meilleur gain de simplicité.

Remarque :

Dans ce travail, nous avons développé une approche permettant l'extraction des règles à partir des lignes du tableau de contingence. Or l'utilisation des approches précédentes conduit également à l'extraction des règles à partir des colonnes d'un tel tableau. L'ensemble de ces règles (lignes et colonnes) constituent le modèle global du système. Leur utilisation doit faire appel à un moteur d'inférence qui génère la ou les règle(s) à appliquer pour la résolution d'un problème ou fait donné. Un tel moteur peut aussi bien combiner des règles lignes (chainage avant ou mode progressif) ou des règles colonnes (chainage arrière ou mode régressif) ou les deux à la fois pour affiner ou confirmer une solution donnée.

VI.5 Modélisation de Y : explication d'une colonne par l'ensemble des lignes contrastées

Soit $[r_{ij}]$ le tableau correspondant à l'ensemble des lignes contrastées. L'opération de contraste a fait apparaître pour chaque ligne de ce tableau une règle de la forme :

$$X = \alpha \text{ ALORS } Y \in B_\sigma(\alpha)$$

ou encore

$$X = \alpha \text{ ALORS } Y \notin M_Y \setminus B_\sigma(\alpha)$$

Le regroupement des lignes et la simplification résultante se sont alors traduites par une modélisation du type :

$$X \in A_i \text{ ALORS } Y \in B_i$$

ou

$$X \in A_i \text{ ALORS } Y \notin M_Y \setminus B_i$$

En d'autres termes, si on note y une variable dichotomique telle que :

$$Y \in B_i \Leftrightarrow y = 1$$

$$Y \notin B_i \Leftrightarrow y = 0$$

On peut exprimer le tableau de contingence contrasté sous la forme suivante :

x \ y	0	1
A _i	0	1
\bar{A}_i	/	/

\bar{A}_i étant le complément de A_i

Tableau IV.7

Ce modèle est parfaitement cohérent avec l'expression des règles de type 1, l'information supplémentaire apportée par les règles de type 2 se traduit alors par la superposition au modèle précédent d'une distribution de probabilité sur chaque ensemble B_i .

Cependant, on peut utiliser le tableau contrasté pour tester une explication locale de Y , modalité par modalité. Par exemple, si l'on cherche à modéliser la modalité $Y = \beta_j$, on définira une variable dichotomique y_j telle que :

$$y_j = 1 \Leftrightarrow Y = \beta_j$$

$$y_j = 0 \Leftrightarrow Y \neq \beta_j$$

Le tableau de contingence devient alors un tableau à deux colonnes :

X \ y _j	0	1
A _j ⁻	1	0
A _j		

Tableau IV.8

La colonne y_j = 1 n'est autre que la colonne β_j du tableau [r_{ij}] qui s'écrit :

$$R(Y=\beta_j / X) = \left\{ r_{j/i}, i = 1, \dots, n \quad \text{tel que : } r_{j/i} = \frac{P_{j/i}}{\sum_1^n P_{j/i}} \right\} \quad (\text{IV.36})$$

Comme le montre la figure ci-dessus, cette colonne partitionne l'ensemble des lignes en deux classes :

$$A_j^- = \{\alpha_i, i = 1, \dots, n \text{ tel que : } r_{j/i} = 0\}$$

$$A_j = \{\alpha_i, i = 1, \dots, n \text{ tel que : } r_{j/i} \neq 0\}$$

$$\text{avec } A_j^- \cup A_j = M_X, j = 1, \dots, m$$

La modélisation de la modalité β_j qui en découle est :

$$X = \alpha_i, \alpha_i \in A_j^- \text{ ALORS } Y \neq \beta_j$$

$$X = \alpha_i, \alpha_i \in A_j \text{ ALORS } Y = (\beta_j, r_{j/i})$$

Une écriture globale de ce modèle sous la forme :

$$X = \bigvee_{\alpha_i \in A_j} \bigwedge_{l=1}^q \alpha_i^l \text{ ALORS } Y = \beta_j \quad (\text{IV.37})$$

n'est pas envisageable, puisqu'elle ne tient pas compte des degrés de certitude différents associés à chacun des couples (α_i, β_j), α_i ∈ A_j. Une approche "floue" conduirait à une écriture globale de la forme :

$$X = \bigvee_{\alpha_i \in A_j} \bigwedge_{i=1}^q \alpha_i^1 \quad \text{ALORS} \quad Y = (\beta_j, r_j^*) \quad (\text{IV.38})$$

$$\text{avec } r_j^* = \min_{\alpha_i \in A_j} r_{ji}$$

L'approche consistant à modéliser une seule modalité de Y revient à ne pas utiliser l'information contenue dans la colonne $y_j = 0$ du tableau IV.8. Une faible valeur de r_j^* traduit le fait que le coefficient de vérité de la règle (IV.38) est faible, l'utilité de l'information perdue pouvant être élevée. La notion de modèle local développée dans la première partie de ce travail peut alors être utilisée : en notant ν un niveau limite de probabilité conditionnelle, on définit alors une variable trichotomique z_j en scindant l'ensemble A_j en deux classes :

$$A_j^0(\nu) = \{ \alpha_i, i = 1, \dots, n \text{ tel que : } r_{ji} < \nu \}$$

$$A_j^1(\nu) = \{ \alpha_i, i = 1, \dots, n \text{ tel que : } r_{ji} \geq \nu \}$$

Ainsi z_j prendra trois modalités β_j^1, β_j^0 et β_j , le tableau IV.9 ci-dessous exprime cette variable.

X \ z _j	β_j^1	β_j^0	$\overline{\beta_j}$
A_j^-	0	0	1
A_j^0	0	1	0
A_j^1	1	0	0

$$\overline{\beta_j} = \{ M_y \setminus \beta_j \}$$

Tableau IV.9

Le modèle de $Y = \beta_j$ s'exprime alors sous la forme suivante :

$$X = \bigvee_{\alpha_i \in A_j^1} \bigwedge_{i=1}^p \alpha_i^1 \quad \text{ALORS} \quad Y = (\beta_j, \nu) \quad (\text{IV.39})$$

$$X = \bigvee_{\alpha_i \in A_j^0} \bigwedge_{i=1}^p \alpha_i^1 \quad \text{ALORS} \quad Y = ?$$

$$X = \bigvee_{\alpha_i \in A_j^-} \bigwedge_{i=1}^p \alpha_i^1 \quad \text{ALORS} \quad Y = \overline{\beta_j}$$

X se présente alors comme une fonction multivaluée qu'il est possible de simplifier. L'utilisation de l'entropie du Max conjointement avec l'entropie de SHANNON conduit à la simplification du modèle. Par exemple, si on note S un monôme avec $S \in \mathcal{P}(X)$:

$$H_M(z/S) = 0, H_M(z/SX) \neq 0 \text{ et } H_S(Y/S) \neq 0$$

signifient que S est l'un des monômes du modèle ; celui-ci n'est pas unique puisque l'entropie de SHANNON est encore non nulle.

Par ailleurs, à chaque niveau ν on peut associer un modèle de la modalité $Y = \beta_j$. En effet, la classe A_j contient un nombre de lignes constant d'où :

$$\text{card} [A_j^1(\nu) \cup A_j^0(\nu)] = \text{constante}$$

Par ailleurs, pour $\nu < \nu'$ on a :

$$A_j^1(\nu) \supset A_j^1(\nu') :$$

Le nombre de modalités de X entraînant $Y = \beta_j$ avec une probabilité $\geq \nu$ augmente lorsque ν diminue. On ne peut pas à priori savoir si cette augmentation se traduira par une simplification de la partie hypothèse.

Réciproquement, pour une partie hypothèse donnée, il est possible de rechercher le niveau ν auquel elle conduit, ce niveau pouvant par ailleurs être acceptable ou non.

VII - APPLICATION : AIDE A LA DECISION AU DIAGNOSTIC MEDICAL

VII.1 Position du problème : BARBOUCHA [1987], AYGALINC [1986] STAROSWIECKI [1986]

On considère l'ensemble des nouveaux-nés admis au centre de Pathologie Néo-Natale d'Arras et atteints de troubles respiratoires. Pour chaque prématuré $\omega_i \in \Omega$ (population d'apprentissage de taille L), un dossier médical est constitué et comprend :

- $X(\omega_i)$; $i \in \{1, \dots, n\}$: les données cliniques relatives à l'historique de l'enfant avant son entrée au centre (âge gestationnel, évolution de la grossesse, accouchement, ...) et les données paracliniques relevées durant son séjour au centre (résultats d'examens biologiques, radiologiques, ...)
- $Y(\omega_i)$; $i \in \{1, \dots, n\}$: l'ensemble des diagnostics qui résultent de l'analyse et de la

synthèse des différentes données recueillies sur le prématuré ω_i .

Devant l'extrême abondance des données qui concernent chaque malade, il est important d'extraire celles qui décrivent de façon satisfaisante telle entité diagnostic ou pronostic. En effet, l'objectif visé par une approche d'aide à la décision est de faciliter le travail du médecin en lui proposant compte tenu des données caractérisant le nouveau-né admis au centre, le ou les diagnostics les plus probables ou les plus efficaces appartenant au champ de la pathologie associée au symptôme d'entrée. La mise en œuvre de l'approche proposée, conduit à extraire des règles à caractère directement opérationnel à partir de la connaissance contenue dans la population d'apprentissage Ω . Cependant, sous son aspect utilisation une méthode d'aide à la décision dans le domaine du diagnostic ne peut se baser à un instant donné que sur l'exploitation des connaissances contenues dans Ω avant cette date.

Dans le cadre d'une médecine d'urgence, on doit émettre des propositions de diagnostic à tout instant et particulièrement dès les premières heures de vie du prématuré pour lequel on ne dispose guère d'information suffisante. Il est par conséquent nécessaire de savoir, quel ensemble de règles faut-il appliquer à l'instant t compte tenu des données dont on dispose à l'instant $(t-1)$ et quels examens complémentaires faut-il prescrire pour affiner ou confirmer les règles à appliquer ou le diagnostic à proposer à l'instant $t+1$. Ceci nécessite la mise en œuvre d'un processus d'exploitation des règles pour l'aide à la décision diagnostique. En effet, on a d'après la population d'apprentissage un ensemble de règles atomiques ou simplifiées. A l'instant t , on ne connaît qu'une partie des variables. Ceci se traduit par un regroupement des règles contenant dans la partie hypothèse ce sous-ensemble de variables mesuré. L'efficacité de la règle composée ainsi obtenue est alors plus faible que l'efficacité des règles initiales. Le recours à des variables explicatives supplémentaires permet d'éliminer un certain nombre de règles parmi celles retenues à l'instant t . A ce niveau, deux démarches peuvent être envisagées :

- problème d'affinement (approche progressive) : utilisation des règles ligne
- problème de confirmation (approche régressive) : utilisation des règles colonne.

VII.2 Mise en œuvre du processus d'aide à la décision

VII.2.1 Introduction

Nous avons montré dans l'approche proposée, qu'il est possible de générer un modèle simplifié du système à partir du regroupement des règles atomiques et de la simplification qui en résulte. Moyennant une renumérotation donnée, soit $P_K(M_X) = \{A_i, i = 1, \dots, K\}$ la partition de M_X à partir de laquelle K règles, écrites sous forme simplifiée sont extraites. A chaque classe A_i , on associe une modalité $\alpha_i \in M_X$ à laquelle on fait correspondre B_i la partie conclusion de la règle. L'ensemble des règles simplifiées extraites des lignes du tableau de contingence s'écrit :

$$X = \alpha_i \text{ ALORS } Y = B_i \quad \forall i \in \{1, \dots, K\}$$

Les règles qui caractérisent à l'instant t-1 la population d'apprentissage sont connues, il convient alors de les utiliser à l'instant t.

Soit $S(\omega_{L+1}) = \alpha$, $S \in \mathcal{P}(X)$, l'ensemble des données caractérisant le prématuré ω_{L+1} et pour lequel une aide au diagnostic est désirée à l'instant t et $\alpha \in M_S$ la modalité prise par $S(M_S \subseteq M_X)$. Cependant, la quantité d'information contenue dans les données recueillies sur le prématuré durant ses premières heures de vie est généralement insuffisante et le médecin est amené à prendre une décision thérapeutique sans avoir acquis une complète certitude au niveau du diagnostic. En effet, plusieurs situations peuvent se produire.

a) $S(\omega_{L+1}) = \alpha \in M_X$: la modalité α est contenue dans la base des règles, il s'agit alors d'appliquer à l'individu ω_{L+1} la règle correspondante. Autrement dit, une telle opération consiste à proposer au médecin la conclusion de la règle qui peut être aussi bien unique ($\sigma = 1$) que multiple ($\sigma > 1$).

b) $S(\omega_{L+1}) = \alpha \in \mathcal{P}(M_X)$: la modalité α est contenue dans la partie hypothèse de plusieurs règles. Ceci revient en réalité à extraire à partir de la population Ω le sous-échantillon Ω_α constitué de l'ensemble des individus identiques à ω_{L+1} . Cette opération conduit naturellement à appliquer l'ensemble des règles contenant $S = \alpha$ dans la partie hypothèse et à proposer au médecin les conclusions tirées de cet ensemble de règles. Dans une telle situation, les conclusions tirées sont forcément multiples puisque les règles du modèle sont écrites sous forme simplifiées, chacune d'elles possède sa propre conclusion.

c) $S(\omega_{L+1}) = \alpha$, tel que α n'appartient pas à une seule partie hypothèse des règles mais peut être reconstitué par l'utilisation d'un ensemble d'hypothèses des règles. Une telle situation se produit lorsque la taille de la population d'apprentissage est telle que :

$$\text{card}(M_X) = n \leq M_{X1} \cdot M_{X2} \cdot \dots \cdot M_{Xp}$$

Supposons que $\alpha = \alpha'_i \wedge \alpha''_k$ avec $\alpha_i = \alpha'_i \wedge \alpha''_i$ et $\alpha_k = \alpha'_k \wedge \alpha''_k$

La règle à appliquer à l'individu ω_{L+1} dans ce cas n'est autre que l'intersection des règles contenant α_i et α_k dans leur partie hypothèse.

En effet : $X = \alpha_i \Rightarrow Y \in B_i$ ou bien $Y \notin B_i = B(\alpha_i) \setminus B_i$
 $X = \alpha_k \Rightarrow Y \in B_k$ ou bien $Y \notin B_k = B(\alpha_k) \setminus B_k$
 $X = \alpha = \alpha'_i \wedge \alpha''_k \Rightarrow Y \notin B_i \wedge Y \notin B_k \Rightarrow Y \in B_i \cap B_k$

L'intersection de B_i et de B_k ne peut être vide puisque α est obtenue par une combinaison de α_i et α_k .

Dans le cas où α est reconstitué par plusieurs α_i , $i \in (1, \dots, k)$, la règle à appliquer est alors obtenue par l'intersection de l'ensemble de règles mises en jeu et la conclusion peut être comme dans le premier cas unique ou multiple.

Selon la nature des conclusions proposées, le médecin peut être amené à adopter différents comportements.

VII.2.2 Les différents comportements du médecin

En se basant sur la proposition diagnostique formulée par les règles à l'instant t , le médecin peut adopter différents comportements à l'instant $t+1$.

- Le diagnostic à l'instant t , noté Y_t , est unique et correspond à l'application de l'une ou l'intersection de plusieurs règles du modèle proposé. Il s'agit pour le praticien de garantir à l'instant $(t+1)$ ce résultat. Si le diagnostic à cet instant confirme un tel résultat, l'objectif escompté est alors atteint, le processus d'aide à la décision doit s'arrêter à moins que d'autres confirmations ne soient sollicitées par le médecin.

- Le diagnostic à l'instant t est multiple et résulte de l'application d'une ou plusieurs règles. Dans les deux cas, il s'agit d'un problème d'affinement du résultat par la mise en œuvre d'examens complémentaires. Ainsi, de proche en proche le processus de diagnostic converge vers une solution unique et l'objectif escompté peut alors être atteint. Si par contre à l'instant $t+1$, les objectifs escomptés ne sont pas atteints, trois types de situation peuvent se produire.

i) $Y_{t+1} \neq Y_t$: la proposition diagnostique à l'instant $t+1$ n'est pas cohérente avec celle émise à l'instant t . Comme cette incohérence n'apparaît qu'à l'instant $t+1$, on peut penser alors que cette situation résulte d'une saisie incorrecte de S sur ω_{L+1} . Dans ce sens, le processus d'aide à la décision propose au médecin de vérifier l'acquisition des dernières mesures. Néanmoins, si cette incohérence s'avère réelle, le processus se trouve inefficace, seul le médecin est en mesure de débloquer la situation, soit en remettant en cause les diagnostics antérieurs, son propre diagnostic ou bien en formulant d'autres propositions.

ii) Y_{t+1} n'affine pas Y_t le résultat acquis à l'instant t . Ceci peut s'expliquer par le fait que les examens complémentaires n'ont pas apporté une quantité d'information suffisante susceptible d'affiner la décision antérieure. Autrement dit, les examens effectués sont en partie redondants. Dans ces conditions, le médecin peut proposer un autre univers de choix de variables explicatives pour remédier à cette situation.

iii) Le cardinal de Ω_α est trop faible pour qu'une proposition diagnostique puisse être émise dans de bonnes conditions [PICARD 72] [TERRENOIRE 76] [ZIGHED 84]. Cela correspond à la règle : $S = \alpha$ ALORS $Y \in B_\sigma(\alpha)$ qui se base sur un nombre très restreint de cas (r_i , faible). Une telle règle est alors peu fiable. Cette situation peut résulter également des erreurs émises dans les diagnostics antérieurs ayant conduit à classer à tort l'individu ω_{L+1} dans le sous-échantillon Ω_α ou bien un tel individu a peu de voisin dans Ω et il est possible qu'il constitue un cas particulier nécessitant le recours au praticien.

VII.2.3 Aspect conversationnel : analyse des différentes situations

a) Les objectifs escomptés sont atteints

En effet, si à l'instant $t+1$ la proposition diagnostique $Y_{t+1} = Y_t$, le médecin peut alors prendre une décision thérapeutique. Le processus d'aide à la décision doit s'arrêter à moins que d'autres confirmations ne soient sollicitées par le médecin.

b) Les objectifs escomptés ne sont pas atteints

Dans une telle situation, le recours à des examens complémentaires s'avère indispensable pour affiner la décision Y_t . Le problème qui se pose au médecin est celui du choix de leur ordre de réalisation et de l'opportunité de poursuivre les investigations au vu des résultats des examens déjà réalisés [LE MINOR 80]. Or dans l'approche proposée, nous avons montré qu'il est possible d'extraire un ensemble de règles à partir des colonnes du tableau de contingence. Le médecin peut alors émettre une proposition diagnostique et demande au système d'aide de formuler l'ensemble des examens complémentaires non encore effectués sur l'individu ω_{L+1} . L'application des règles colonnes permet au processus d'aide de proposer directement l'ensemble des examens complémentaires et leur ordre de réalisation respectif. Connaissant les délais d'obtention des résultats qui peuvent varier de quelques heures à quelques jours, leur coût financier, le risque de complication, ..., le médecin peut être amené à modifier l'ordre de ces examens complémentaires. Après une telle opération, si les objectifs escomptés sont atteints, le processus d'aide s'arrête. Dans le cas contraire, le médecin peut émettre une 2^{ème} proposition, et ainsi de suite jusqu'à ce que le processus d'aide converge vers un diagnostic permettant au médecin de prendre une décision définitive.

D'une façon générale, l'approche proposée permet la décomposition du problème de décision dynamique qu'est une aide de diagnostic dans un contexte de médecine d'intervention en une suite de problèmes de décision statique. L'ensemble des décisions admissibles à l'instant $t+1$ sont choisies en fonction des règles proposées par le système d'aide à l'instant t . La décision la mieux appropriée dépend :

- ☛ de la démarche adoptée par le praticien et des règles reconnues valides à l'instant t par le processus d'aide à la décision ;
- ☛ du poids et de la fiabilité de l'information, des contraintes liées au délai, à la souffrance des malades, au risque de complication de la maladie, à la dangerosité, et au coût financier que nécessitent les examens complémentaires pour l'affinement ou la confirmation d'un résultat donné.

VIII - CONCLUSION

En l'absence du modèle du système, nous avons proposé une approche permettant à partir des données recueillies sur celui-ci, la génération d'un ensemble de règles atomiques. Celles-ci constituent le modèle provisoire du système. Après l'opération de regroupement des règles et la simplification qui en résulte, nous avons montré qu'il était possible d'établir le modèle simplifié du système qui peut être affiné ou mis à jour à chaque fois que de nouvelles données sont disponibles. L'utilisation d'une telle approche qui opère à partir des tableaux contrastés (opérateurs C_S o RG et RG o C_G) peut faciliter la représentation et l'extraction des connaissances contenues dans les observations recueillies sur le système. Par ailleurs, l'expert du domaine peut formuler les règles qui régissent le fonctionnement du système. Celles-ci conceptualisent son expérience professionnelle et son savoir faire acquis sur le long terme. Il convient alors de comparer ces règles avec celles générées par l'approche proposée. Ceci revient en réalité à tester si les échantillons relevés sur le système et les règles auxquelles ils donnent naissance sont ou non en accord avec les règles données par l'expert. Deux situations peuvent se produire :

- ☛ Les règles sont cohérentes : l'expert confirme les résultats des échantillons analysés ;
- ☛ Les règles sont incohérentes : une telle situation peut résulter d'une saisie incorrecte d'un ensemble de variables et il convient alors de vérifier leur validité. Dans le cas où celle-ci s'avère correcte, seul l'expert est en mesure de résoudre le problème en faisant appel à des analyses supplémentaires pour affiner ou confirmer le résultat ou tout simplement en remettant en cause ses propres conclusions.

En définitive, l'approche proposée constitue un extracteur de connaissance pouvant être utilisé pour la vérification ou la mise à jour des règles d'un système expert et peut aussi être appliqué aux systèmes technologiques soit pour la détection des différents modes de fonctionnement (normal, défaillant, ...) soit pour le diagnostic des pannes.

CONCLUSION GENERALE

L'analyse d'un système complexe sur lequel on ne dispose d'aucune information à priori passe forcément par l'analyse des données qu'il génère. Il convient dans ce sens de fournir à l'analyste des outils et des méthodes facilitant son approche vers la modélisation du système, par l'extraction et la structuration des connaissances contenues dans ces données.

Dans ce travail, nous avons abordé deux aspects de la modélisation, d'une part l'approche analyse structurale et d'autre part l'approche système expert. Dans une première partie, nous avons montré que les méthodes géométriques de l'analyse de données peuvent être utilisées. Or chaque méthode possède son domaine privilégié d'application et sa propre spécificité (nature des données, l'objectif visé par l'analyse ...). La plupart de ces méthodes utilisent l'hypothèse de linéarité du modèle représentatif du système et ne peuvent par conséquent mettre en évidence ses caractéristiques structurelles. C'est pourquoi, nous avons fait appel à la théorie de l'information qui permet de déceler les relations aussi bien linéaires que non linéaires, qui s'applique à tous types de variables : quantitatives ou qualitatives et n'exige aucune hypothèse restrictive. L'utilisation d'une telle approche nous a conduit à proposer des résultats en modélisation structurale. Plusieurs aspects importants ont été mis en relief : les concepts de modélisabilité, sous-ensembles explicatifs uniques ou multiples d'une variable, degrés de liberté du système, les relations redondantes, recherche de variables explicatives supplémentaires ... Nous avons présenté à cet effet deux algorithmes permettant de faciliter la procédure de modélisation proprement dite. Nous avons souligné les inconvénients d'une approche globale du système qui travaille sur l'ensemble des échantillons relevés sur celui-ci et avons montré l'intérêt des approches locales qui utilisent des échantillons de taille plus réduite.

Cette façon de procéder nous a conduit à aborder un deuxième aspect de la modélisation : l'extraction de connaissances sous forme de règles à partir d'un tableau de contingence. Dans ce cas, il s'agit d'une modélisation "experte" qui consiste à extraire les connaissances que recèle la population d'apprentissage représentative du système analysé. Dans cette deuxième phase de modélisation, deux approches similaires sont développées, la première est basée sur la définition d'un opérateur de contraste, le deuxième utilise un coefficient d'extraction de l'information. Ceux-ci constituent des extracteurs de connaissance et génèrent les règles qui régissent le fonctionnement du système. Ces règles forment en fait le modèle provisoire du système qu'il convient de simplifier. Nous avons montré que l'entropie du Max peut, dans certains cas, aboutir à une telle simplification et le modèle simplifié du système donné sous forme de règles, peut

alors être établi. Le ou les experts du domaine analysé proposent les règles qui sont censées décrire le comportement d'une population d'apprentissage, celles-ci sont à comparer avec les règles proposées par l'approche que nous avons développée. Une procédure de mise à jour des règles est alors nécessaire si les règles formulées dans les deux ne s'avèrent pas cohérentes.

Nous avons développé une application dans le domaine de l'aide à la décision médicale, et avons montré que les règles extraites de la population d'apprentissage peuvent être directement opérationnelles. Cette approche peut de la même façon s'appliquer dans d'autres domaines : classification automatique, diagnostic technologique, etc ...

Loin d'être épuisé, le domaine de modélisation que nous avons voulu explorer dans cette thèse nécessite encore bien des recherches. Outre la nécessité de l'expérimentation et la mise en œuvre pratique des approches proposées, on peut discerner quelques directions de recherches qui nous semblent intéressantes.

En premier lieu, développer de nouveaux indices pour juger la faisabilité du problème de décomposition du système en sous-systèmes faiblement couplés. De tels indices doivent tenir compte du bruit qui peut se superposer à la fois aux lignes et aux colonnes du tableau de contingence. Ainsi, la définition d'un indice "informationnel" calculé à partir du tableau de contingence contrasté horizontalement et verticalement permettra d'apprécier plus efficacement les structures de liaison entre variables et groupes de variables du système. Par ailleurs, le critère généralement utilisé pour la décomposition est celui qui aboutit à des sous-systèmes faiblement couplés. Or, il ne serait pas sans intérêt de faire appel à une analyse multi-critère faisant intervenir les notions de complexité des sous-systèmes, de coût, de fiabilité, ... etc.

Dans un deuxième temps, développer des méthodes d'identification basées sur les concepts entropiques. Nous avons montré que l'utilisation de l'entropie de SHANNON conjointement avec l'entropie du Max permet l'identification des monômes d'une fonction logique combinatoire. Il serait intéressant d'étendre cette approche à d'autres types de systèmes, les systèmes séquentiels par exemple, ... etc.

Enfin, il serait souhaitable de développer un outil informatique, permettant la mise en œuvre de l'approche extraction des connaissances pour la vérification et la mise à jour des règles des systèmes experts.

ANNEXES

ANNEXE II

Y \ X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	2	0	1	20	1	1	2	0	1	2	0	0	0	0	0	0	30
2	1	1	25	0	1	2	1	2	1	1	0	0	0	0	0	0	35
3	1	1	0	1	25	1	1	2	1	0	0	0	0	0	0	0	33
4	0	0	1	1	0	24	0	0	0	0	0	0	0	0	0	0	26
5	0	1	0	1	0	0	20	1	1	1	0	0	0	0	0	0	25
6	0	0	0	1	1	15	1	1	1	1	0	0	0	0	0	0	21
7	1	1	0	0	15	1	1	0	1	0	0	0	0	0	0	0	20
8	0	0	1	1	0	0	24	1	1	2	0	0	0	0	0	0	30
9	1	24	1	0	0	2	1	0	0	0	0	0	0	0	0	0	29
10	30	1	2	1	0	0	0	2	1	0	0	0	0	0	0	0	37
11	0	1	20	1	2	1	0	0	1	0	0	0	0	0	0	0	26
12	0	1	1	20	0	0	1	1	0	0	0	0	0	0	0	0	24
13	1	25	1	0	1	1	2	0	0	1	0	0	0	0	0	0	32
14	25	1	1	1	0	1	1	0	2	0	0	0	0	0	0	0	32
15	0	0	0	1	1	1	0	25	1	1	0	2	25	0	0	0	57
16	0	0	0	1	1	1	2	2	20	2	25	1	1	0	0	10	66
17	0	0	0	1	1	1	0	1	0	20	2	1	0	1	30	0	58
18	0	0	0	0	0	1	1	0	0	30	1	3	0	0	0	25	61
19	0	0	0	1	1	0	0	0	2	1	1	45	1	30	0	0	82
20	0	0	0	0	1	1	1	0	0	1	35	0	0	1	20	0	60
21	0	0	0	0	1	1	1	0	20	1	1	0	0	35	0	1	61
22	0	0	0	0	1	0	0	1	1	2	1	15	15	0	0	1	37
23	0	0	0	0	1	2	1	20	35	2	2	2	2	0	0	0	65
24	0	0	0	0	1	0	1	1	2	2	1	30	0	0	15	0	53
	62	57	54	52	55	57	62	60	92	70	69	99	42	67	65	37	1000

TABLEAU I : Explication de Y par l'ensemble des variables explicatives primaires (X)



Y \ XUS*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	2	0	1	20	1	1	2	0	1	2	0	0	0	0	0	0	30
2	1	1	25	0	1	2	1	2	1	1	0	0	0	0	0	0	35
3	1	1	0	1	25	1	1	2	1	0	0	0	0	0	0	0	33
4	0	0	1	1	0	24	0	0	0	0	0	0	0	0	0	0	26
5	0	1	0	1	0	0	20	1	1	1	0	0	0	0	0	0	25
6	0	0	0	1	1	15	1	1	1	1	0	0	0	0	0	0	21
7	1	1	0	0	15	1	1	0	1	0	0	0	0	0	0	0	20
8	0	0	1	1	0	0	24	1	1	2	0	0	0	0	0	0	30
9	1	24	1	0	0	2	1	0	0	0	0	0	0	0	0	0	29
10	30	1	2	1	0	0	0	2	1	0	0	0	0	0	0	0	37
11	0	1	20	1	2	1	0	0	1	0	0	0	0	0	0	0	26
12	0	1	1	20	0	0	1	1	0	0	0	0	0	0	0	0	24
13	1	25	1	0	1	1	2	0	0	1	0	0	0	0	0	0	32
14	25	1	1	1	0	1	1	0	2	0	0	0	0	0	0	0	32
15	0	0	0	1	1	1	0	25	0	1	0	1	0	0	0	0	30
16	0	0	0	1	1	0	1	1	20	0	0	0	0	0	0	0	24
17	0	0	0	1	1	1	0	1	0	20	1	1	0	0	0	0	26
18	0	0	0	0	0	1	1	0	0	30	0	2	0	0	0	0	34
19	0	0	0	1	1	0	0	0	1	1	1	45	0	0	0	0	50
20	0	0	0	0	1	1	1	0	0	0	35	0	0	0	0	0	38
21	0	0	0	0	1	1	1	0	20	0	0	0	0	0	0	0	23
22	0	0	0	0	1	0	0	1	1	1	1	15	0	0	0	0	20
23	0	0	0	0	1	1	0	0	35	1	1	1	0	0	0	0	40
24	0	0	0	0	0	1	1	20	0	1	1	1	0	0	0	0	25
25	0	0	0	0	0	1	1	1	0	1	25	1	0	0	0	0	30
26	0	0	0	0	1	0	1	1	1	1	0	30	0	0	0	0	35
27	0	0	0	0	0	0	0	0	1	0	0	1	25	0	0	0	27
28	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	10	12
29	0	0	0	0	0	0	0	0	0	0	1	0	0	1	30	0	32
30	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	25	27
31	0	0	0	0	0	0	0	0	1	0	0	0	1	30	0	0	32
32	0	0	0	0	0	0	0	0	0	1	0	0	0	1	20	0	22
33	0	0	0	0	0	0	0	0	0	1	1	0	0	35	0	1	38
34	0	0	0	0	0	0	0	0	0	1	0	0	15	0	0	1	17
35	0	0	0	0	0	0	0	0	1	1	1	0	0	0	15	0	18
	62	57	54	52	55	57	62	60	92	70	69	99	42	67	65	37	1000

TABLEAU II : Explication de Y par les variables primaires
 et un sous ensemble de variables explicatives
 secondaires



TABLEAU III : Echantillons faiblement concentrés

$X^1 \backslash Y^1$	1	2	3	4	5	6	7	8	9	10	
1	2	0	1	20	1	1	2	0	1	2	30
2	1	1	25	0	1	2	1	2	1	1	35
3	1	1	0	1	25	1	1	2	1	0	33
4	0	0	1	1	0	24	0	0	0	0	26
5	0	1	0	1	0	0	20	1	1	1	25
6	0	0	0	1	1	15	1	1	1	1	21
7	1	1	0	0	15	1	1	0	1	0	20
8	0	0	1	1	0	0	24	1	1	2	30
9	1	24	1	0	0	2	1	0	0	0	29
10	30	1	2	1	0	0	0	2	1	0	37
11	0	1	20	1	2	1	0	0	1	0	26
12	0	1	1	20	0	0	1	1	0	0	24
13	1	25	1	0	1	1	2	0	0	1	32
14	25	1	1	1	0	1	1	0	2	0	32
	62	57	54	48	46	49	55	10	11	8	400= Ω_1



TABIEAU IV : Echantillons moyennement concentrés

$X^2 \backslash Y^2$	4	5	6	7	8	9	10	11	12			
15	1	1	1	0	2	5	0	1	0	1	30	
16	1	1	0	1	1	2	0	0	0	0	24	
17	1	1	1	0	1	0	2	0	1	1	26	
18	0	0	1	1	0	0	3	0	0	2	34	
19	1	1	0	0	0	1	1	1	1	4	50	
20	0	1	1	1	0	0	0	3	5	0	38	
21	0	1	1	1	0	2	0	0	0	0	23	
22	0	1	0	0	1	1	1	1	1	1	20	
23	0	1	1	0	0	3	5	1	1	1	40	
24	0	0	1	1	2	0	0	1	1	1	25	
25	0	0	1	1	1	0	1	2	5	1	30	
26	0	1	0	1	1	1	1	0	3	0	35	
	4	9	8	7	5	0	7	8	5	7	65	97
											375 = Ω_2	

TABIEAU V : Echantillons fortement concentrés

$X^3 \backslash Y^3$	9	10	11	12	13	14	15	16			
27	1	0	0	1	2	5	0	0	0	27	
28	0	1	0	0	1	0	0	0	1	12	
29	0	0	1	0	0	1	3	0	0	32	
30	0	0	1	1	0	0	0	2	5	27	
31	1	0	0	0	1	3	0	0	0	32	
32	0	1	0	0	0	1	2	0	0	22	
33	0	1	1	0	0	3	5	0	1	38	
34	0	1	0	0	1	5	0	0	1	17	
35	1	1	1	0	0	0	1	5	0	18	
	3	5	4	2	4	2	6	7	6	5	37
											225 = Ω_3



ANNEXE III

ANNEXE III.1

Il s'agit de minimiser $D(P,R)$ sous la contrainte $\sum_1^{\sigma} r_{j/i} = 1$. On considère le sous tableau de contingence relatif à la sous population Ω_i . Dans un tel tableau, on considère les deux distributions P et R telles que :

$$P_j = \begin{cases} P_{ij} + r_{ij} & \text{pour } j = 1, \dots, \sigma \\ P_{ij} & \text{pour } j = \sigma+1, \dots, m \end{cases} \quad \text{et } P_i = r_i.$$

III.1.1. Distance du χ^2

$D_1(P, R) = \sum_1^m (1/P_j) (r_{j/i} - P_{j/i})^2$ calculons le Lagrangien :

$$L = \sum_1^m (1/P_j) (r_{j/i} - P_{j/i})^2 + \lambda \left(\sum_1^{\sigma} r_{j/i} - 1 \right)$$

$$(\partial L / \partial r_{j/i}) = (1/P_i) [(r_{j/i} - P_{j/i}) (r_{j/i} + 3 P_{j/i}) / (P_{j/i} + r_{j/i})^2] + \lambda$$

En notant $x = r_{j/i} / P_{j/i}$ et en faisant $\partial L / \partial r_{j/i} = 0$ on obtient une équation du deuxième degré dont la solution positive s'écrit :

$$x = -1 + (2 / \sqrt{1 + \lambda P_i}) \Rightarrow r_{j/i} = P_{j/i} (2 / \sqrt{1 + \lambda P_i}) - 1$$

Sachant que $\sum_1^{\sigma} r_{j/i} = 1$, on aboutit à :

$$r_{j/i} = P_{j/i} / \sum_1^{\sigma} P_{j/i} \quad \forall j \in [1, \dots, \sigma]$$

Remarque : on obtient le même résultat lorsque on utilise l'indice de distance du χ^2 suivant :

$$\chi^2 = L\alpha_i \left[\sum_1^{\sigma} (r_{ij}^2 / r_i \cdot r_j) + \sum_1^m (P_{ij}^2 / P_i \cdot P_j) - 1 \right]$$

III.1.2 L'information discriminante de KULLBACK

$$J(P, R) = D_2(P, R) = \sum_1^m r_{ji} \text{Log}(r_{ji} / P_{ji})$$

$$L = \sum_1^{\sigma} r_{ji} \text{Log}(r_{ji} / P_{ji}) + \lambda (\sum_1^{\sigma} r_{ji} - 1)$$

$$(\partial L / \partial r_{ji}) = \text{Log}(r_{ji} / P_{ji}) + 1 + \lambda$$

Cette dérivée s'annule lorsque $r_{ji} = P_{ji} 2^{-(1+\lambda)}$

Sachant que $\sum_1^{\sigma} r_{ji} = 1$ on aboutit à :

$$r_{ji} = P_{ji} / \sum_1^{\sigma} P_{ji} \quad \forall j \in [1, \dots, \sigma]$$

III.1.3 Le χ^2 de NEYMAN'S

$$\chi^2 = D_3(P, R) = \sum_1^m [(r_{ji} - P_{ji})^2 / P_{ji}]$$

$$L = D_3(P, R) + \lambda (\sum_1^{\sigma} r_{ji} - 1)$$

$$\partial L / \partial r_{ji} = 2 [(r_{ji} - P_{ji}) / P_{ji}] + \lambda$$

La dérivée s'annule lorsque $r_{ji} = P_{ji} (1 - (\lambda/2))$

Sachant que $\sum_1^{\sigma} r_{ji} = 1$ on obtient :

$$r_{ji} = P_{ji} / \sum_1^{\sigma} P_{ji} \quad \forall j \in [1, \dots, \sigma]$$

III.1.4 L'information de SHANNON :

$$I(P : R) = D_4(P, R) = \sum_1^m P_{ij} \text{Log}(P_{ij} / P_i \cdot P_j) + \sum_1^{\sigma} r_{ij} \text{Log}(r_{ij} / r_i \cdot r_j)$$

$$L = D_4(P, R) + \lambda (\sum_1^{\sigma} r_{ji} - 1)$$

$$\partial L / \partial r_{ji} = \text{Log}(r_{ji} / r_{ji} + P_{ji}) + \lambda + 1$$

L'annulation de la dérivée et à l'aide de la contrainte, on aboutit au même résultat que précédemment.

III.1.5 Indice de distance du type quadratique :

$$D_\sigma (R, P) = \sum_1^m |r_{j/i} - P_{j/i}|^\gamma \quad \text{avec } \gamma > 0$$

$$L = D_\sigma (P, R) + \lambda \left(\sum_1^\sigma r_{j/i} - 1 \right)$$

$$\partial L / \partial r_{j/i} = \gamma (r_{j/i} - P_{j/i})^{\gamma-1} + \lambda$$

L'annulation de la dérivée et à l'aide de la contrainte on aboutit à :

$$r_{j/i} = P_{j/i} + (1/\sigma) \cdot \sum_{\alpha=1}^m P_{j/i}, \quad \forall j \in [1, \dots, \sigma]$$

ANNEXE III.2

L'opérateur de contraste quadratique :

Il est défini comme suit :

$$C_\sigma (j) = \begin{cases} P_{j/i} / \sum_1^\sigma P_{j/i}^2 & \text{pour } j = 1, \dots, \sigma \\ 0 & \text{pour } j = \sigma+1, \dots, m \end{cases}$$

La nouvelle distribution $R (Y / X = \alpha_i)$ s'écrit :

$$R (Y / X = \alpha_i) = P (Y / X = \alpha_i) \cdot C_\sigma (j)$$

Les éléments de $P (Y / X = \alpha_i)$ étant ordonnés, celles de $R (Y / X = \alpha_i)$ le sont également et s'écrivent :

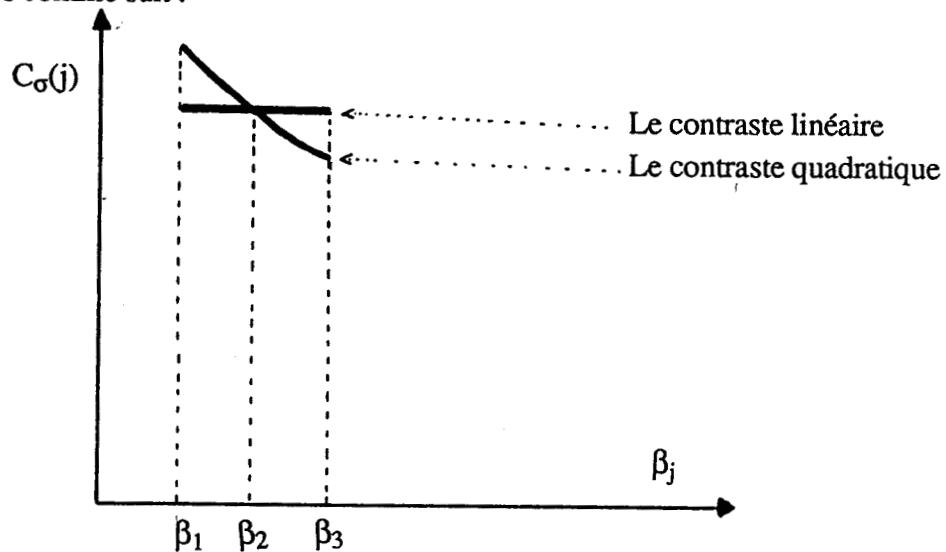
$$R(Y/X=\alpha_i) = \begin{cases} r_{j/i} = P_{j/i} \cdot \left(P_{j/i} / \sum_1^\sigma P_{j/i}^2 \right) = \left(P_{j/i}^2 / \sum_1^\sigma P_{j/i}^2 \right) = \left(P_{ij}^2 / \sum_1^\sigma P_{ij}^2 \right) & \text{pour } j=1, \dots, \sigma \\ r_{j/i} = 0 & \dots \dots \dots \text{pour } j = \sigma+1, \dots, m \end{cases}$$

. Les distributions marginales étant conservées, en effet : $r_{i.} = P_{i.}$

. Il s'agit effectivement d'une distribution de probabilité puisque :

$$\sum_1^{\sigma} r_{j/i} = \sum_1^{\sigma} r_{j/i} = \sum_1^{\sigma} (P_{ij}^2 / \sum P_{ij}^2) = 1$$

Si on prend l'exemple donné par la figure III.4, les deux coefficients de contraste peuvent être schématisés comme suit :



ANNEXE III.3 : LES TESTS STATISTIQUES

III.3.1. Le test du χ^2

Soit $P = \{ P_{j/i}, j = 1, \dots, m \}$ et $R = \{ r_{j/i}, j = 1, \dots, \sigma \} = \{ (P_{ij} / \sum_1^{\sigma} P_{ij}), j = 1, \dots, \sigma \}$

$$\text{avec : } \begin{cases} r_j = P_j = P_{ij} (1 + (P_i / \sum_1^{\sigma} P_{ij})) \text{ pour } j \in \{1, \dots, \sigma\} \\ r_j = 0 \text{ et } P_j = P_{ij} \text{ pour } j \in \{\sigma+1, \dots, m\} \end{cases} \quad r_i = P_i$$

$$\chi^2 / L = \sum_1^{\sigma} (P_{ij}^2 / P_i P_j) + \sum_{\sigma+1}^m (r_{ij}^2 / P_i P_j) + \sum (P_{ij}^2 / P_i P_j) - 1$$

Posons $\sum_1^{\sigma} P_{ij} = S$, il vient :

$$\begin{aligned} &= \sum_1^{\sigma} ((P_{ij}^2 + r_{ij}^2) / P_i P_j) + \sum_{\sigma+1}^m (P_{ij} / P_i) - 1 \\ &= [\sum (P_{ij}^2 + (P_{ij}^2 \cdot P_i^2 / S^2)) / (P_i P_j (1 + P_i / S))] + (P_i - S) / P_i - 1 \\ &= [\sum_1^{\sigma} (P_{ij} (1 - (P_i^2 / S^2))) / (P_i (1 + (P_i / S)))] - (S / P_i) \end{aligned}$$

$$\begin{aligned}
&= [(S^3 + P_i^2) / (S^2 P_i + P_i S)] - (S / P_i) = (P_i - S) / (P_i + S) \\
&= (P_i - \sum_1^{\sigma} P_{ij}) / (P_i + \sum_1^{\sigma} P_{ij}) = (1 - \sum_1^{\sigma} P_{j/i}) / (1 + \sum_1^{\sigma} P_{j/i}) = (\sum_{\sigma+1}^m P_{j/i}) / (1 + \sum_1^{\sigma} P_{j/i})
\end{aligned}$$

C.Q.F.D.

III.3.2 Le test du χ^2 de NEYMAN'S : celui-ci s'écrit :

$$\begin{aligned}
\chi^2 (P, \sigma) &= \sum_1^m ((r_{j/i} - P_{j/i}) / P_{j/i})^2 \quad \text{posons } \sum_1^{\sigma} P_{j/i} = A \\
&= \sum_1^{\sigma} P_{j/i} (1/A - 1)^2 + \sum_{\sigma+1}^m P_{j/i} \\
&= (1/A - 1)^2 \cdot A + (1 - A) \\
&= (1 - A) / A = (1 - \sum_1^{\sigma} P_{j/i}) / \sum_1^{\sigma} (P_{j/i}) = \sum_{\sigma+1}^m P_{j/i} / \sum_1^{\sigma} P_{j/i} \quad \text{C.Q.F.D.}
\end{aligned}$$

Remarquons que $\chi^2 (P, \sigma) = C_{\sigma} - 1$

III.3.3. Test utilisant l'information discriminante de SHANNON

$$I (X : Y) = \sum_{i=1}^n \sum_{j=1}^m P_{ij} \text{Log} (P_{ij} / (P_i \cdot P_j))$$

la perte d'information occasionnée par le contraste de ligne i s'écrit :

$$\begin{aligned}
\Delta I &= \sum_1^m P_{ij} \text{Log} (P_{ij} / (P_i \cdot P_j)) - \sum_1^{\sigma} r_{ij} \text{Log} (r_{ij} / (r_i \cdot r_j)) \\
\text{Posons } A &= \sum_1^{\sigma} P_{j/i} \\
\Delta I &= \sum_1^{\sigma} P_{ij} \text{Log} P_{ij} / [P_i \cdot (P_{ij} + (P_{ij} / A))] + \sum_{\sigma+1}^m P_{ij} \text{Log} (P_{ij} / (P_i \cdot P_j)) \\
&\quad - \sum_1^{\sigma} (P_{ij} / A) \text{Log} (1 / P_i \cdot (1+A)) \\
&= A \cdot P_i \text{Log} (A / (P_i (1+A))) + (P_i - A P_i) \text{Log} (1 / P_i) - P_i \text{Log} (1 / (P_i \cdot (1+A))) \\
&= [A P_i \log A - A P_i \text{Log} P_i - A P_i \text{Log} (1+A)] + [- P_i \text{Log} P_i - A P_i \text{Log} P_i] \\
&\quad + [P_i \text{Log} P_i + P_i \text{Log} (1+A)] \\
&= P_i [A \text{Log} A + (1-A) \text{Log} (1+A)] \quad \text{or } A = \sum_1^{\sigma} P_{j/i} = 1 / C_{\sigma}
\end{aligned}$$

$$= P_i \cdot [(1/C_\sigma) \text{Log} (1/C_\sigma) + (1 - (1/C_\sigma)) \text{Log} (1 + (1/C_\sigma))] \quad \text{C.Q.F.D.}$$

III.3.4. Test de WILCOXON-MANN-WHITNEY

Soient P et R les deux distributions à tester dont les éléments sont supposés ordonnés (ordre croissant)

$$P = \{ P_{j/i}, j = 1, \dots, m \text{ tel que : } P_{m/i} \leq \dots \leq P_{1/i} \}$$

$$\text{et } R = \{ r_{j/i}, j = 1, \dots, \sigma \text{ tel que : } r_{\sigma/i} \leq \dots \leq r_{1/i} \}$$

L'idée du test consiste à effectuer un mélange des deux distributions, les ordonner par valeurs croissantes et vérifier si le mélange (P, R) obtenu est homogène ou non. Si on note U le nombre de couples $(r_{j/i}, P_{k/i})$ pour lesquels $r_{j/i} > P_{k/i}$, on montre alors facilement que U est compris entre deux valeurs extrêmes $U_{\min} = U_m$ et $U_{\max} = U_M$ lorsque les mélanges s'organisent comme suit :

- a. si $(P, R) = \{ 0, \dots, 0, (P_{m/i}, \dots, P_{1/i}), (r_{\sigma/i}, \dots, r_{1/i}) \}$ alors : $U_M = \sigma \cdot m$
- b. si $(P, R) = \{ 0, \dots, 0, P_{m/i}, \dots, P_{\sigma-1/i}, (P_{\sigma/i}, r_{\sigma/i}), \dots, (P_{1/i}, r_{1/i}) \}$ alors

$$U_m = \sigma \cdot m - (\sigma(\sigma - 1)) / 2$$

. Le cas (a) se présente lorsque les résidus sont de faible amplitude et en nombre important, c'est-à-dire $(m - \sigma)$ élevé.

. Le cas (b) s'obtient lorsque les résidus sont faibles en amplitude et en nombre réduit,

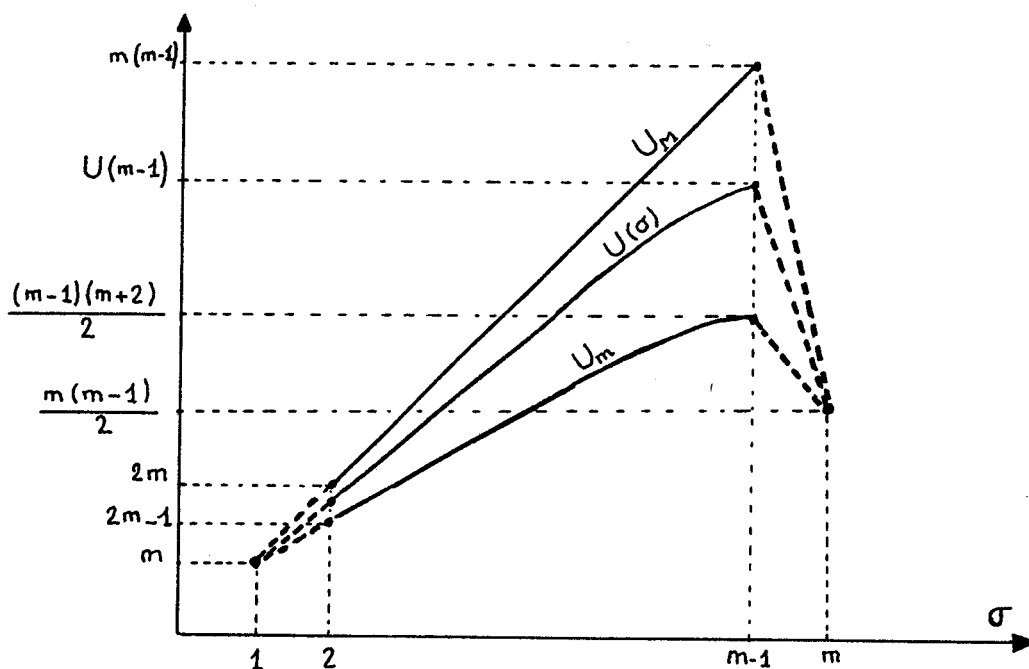
c'est-à-dire $\sum_{\sigma+1}^m P_{j/i}$ petit et $(m - \sigma)$ faible.

La comparaison de U par rapport à U_m et U_M permet de juger si les deux distributions sont proches ou non.

U est connu sous le nom de la statistique de MANN-WHITNEY. Elle est tabulée et permet alors de juger si l'hypothèse H_0 peut être acceptée ou non.

Afin de comparer ce test avec les tests précédents, nous prendrons comme indice de distance $\Delta U (\sigma) = 1/U_{\min} - 1/U$

Pour avoir une idée complémentaire sur ce test, nous avons tracé ci-dessous U , U_M et U_m en fonction σ (figure ci-dessous).



Tracé de $U(\sigma)$, U_M , U_m en fonction de σ

Remarque : Ce test n'apporte aucune information au sujet des deux distributions lorsque $\sigma = m$

III.3.5 Test utilisant la divergence de KULLBACK

La notion de divergence entre lois de probabilités a été introduite par KULLBACK [1962] à partir de l'information de SHANNON. Si on note $P = \{P_{j/i}, j = 1, \dots, m\}$ et $R = \{r_{j/i}, j = 1, \dots, \sigma\}$ les deux distributions de probabilité associées à Y conditionné par $X = \alpha_i$, la divergence de KULLBACK peut être définie comme une fonction qui s'applique aux deux distributions et s'écrit :

$$J \left[\begin{matrix} P \\ R \end{matrix} \right] = \sum_1^m r_{ij} \text{Log} (r_{j/i} / P_{j/i})$$

$$J \left[\begin{matrix} P \\ R \end{matrix} \right] = \sum_1^{\sigma} r_{ij} \text{Log} (r_{j/i} / P_{j/i}) + \sum_{\sigma+1}^m r_{ij} \text{Log} (r_{j/i} / P_{j/i})$$

$$= \sum_1^{\sigma} (P_{ij} / \sum_1^{\sigma} P_{j/i}) \text{Log} (1 / \sum_1^{\sigma} P_{j/i})$$

$$= -P_i \cdot \text{Log} \sum_1^{\sigma} P_{j/i} \quad \text{C.Q.F.D.}$$

$$J \left[\begin{matrix} P \\ R \end{matrix} \right] \text{ est } \geq 0 \text{ puisque } \sum_1^{\sigma} P_{j/i} \leq 1 \Rightarrow -\text{Log} \sum_1^{\sigma} P_{j/i} \geq 0$$

ANNEXE III.4

III.4.1 $U_1(\sigma)$ vérifie la première propriété :

En effet : . si $H(Q) = H_M(Q) = \text{Log}(m - \sigma)$ alors $U_1(\sigma) = U_{1m}(\sigma) = 0$

. si $H(Q) = H_m(Q) = 0$ alors $U_1(\sigma) = U_{1M}(\sigma) = A$

Par ailleurs, si $H(Q_i) \leq H(Q_k)$ avec Q_i et Q_k deux distributions résiduelles et $U_1(Q_i)$, $U_1(Q_k)$ leurs utilités respectives, alors :

$U_1(Q_i) \geq U_1(Q_k)$. Ceci est évident puisque :

$$H(Q_i) \leq H(Q_k) \Rightarrow H(Q_i) / \text{Log}(m - \sigma) \leq H(Q_k) / \text{Log}(m - \sigma) \quad \text{pour } (m - \sigma) > 1$$

$$\Rightarrow 1 - (H(Q_i) / \text{Log}(m - \sigma)) \geq 1 - (H(Q_k) / \text{Log}(m - \sigma))$$

$$\Rightarrow A [1 - (H(Q_i) / \text{Log}(m - \sigma))] \geq A [1 - (H(Q_k) / \text{Log}(m - \sigma))]$$

$$\Rightarrow U_1(Q_i) \geq U_1(Q_k)$$

III.4.2. $U_1(\sigma)$ ne vérifie pas toujours la deuxième propriété

Si on note $H_\sigma(Q) = H(\sigma)$, $H_{\sigma+1}(Q) = H(\sigma+1)$ les entropies résiduelles respectivement de niveau σ et $\sigma+1$ et $U_1(\sigma)$, $U_1(\sigma+1)$, leurs utilités, on montre alors que $U_1(\sigma) - U_1(\sigma+1) = \Delta U_1$ peut être positif ou négatif selon les valeurs de $H(\sigma)$ et $H(\sigma+1)$. En effet, on note :

$$B = \sum_{\sigma+2}^m P_{j/i} = A - P_{\sigma+1/i} \text{ et } \Delta U \text{ s'écrit alors :}$$

$$\Delta U_1 = A [1 - (H(\sigma) / \text{Log}(m - \sigma))] - B [1 - (H(\sigma+1) / \text{Log}(m - \sigma - 1))]$$

$$\Delta U_1 = (A - B) + [B H(\sigma+1) / \text{Log}(m - \sigma - 1)] - [A H(\sigma) / \text{Log}(m - \sigma)] \text{ or}$$

$$A H(\sigma) = B H(\sigma+1) + B \text{Log} [1 + (P_{\sigma+1/i} / B)] + P_{\sigma+1/i} \text{Log} [1 + (B / P_{\sigma+1/i})] > B H(\sigma+1)$$

d'où :

$$\Delta U_1 = P_{\sigma+1/i} [1 - (\text{Log} (1 + (B / P_{\sigma+1/i}))) / \text{Log} (m - \sigma)]$$

$$\underbrace{\hspace{15em}}_{(1)}$$

$$- B [-H(\sigma+1) ((1 / \text{Log} (m - \sigma - 1)) - (1 / \text{Log} (m - \sigma))) + (\text{Log} (1 + (P_{\sigma+1/i} / B))) / \text{Log} (m - \sigma)]$$

$$\underbrace{\hspace{15em}}_{(2)}$$

. si (1) \geq (2) alors : $\Delta U_1 \geq 0$ Par contre

. si (1) \leq (2) alors : $\Delta U_1 \leq 0$

L'exemple ci-dessous confirme que $U_1 (\sigma)$ n'est pas une fonction décroissante de σ .

Exemple : Soit la distribution de probabilité suivante :

$$P (Y / X = \alpha_i) = \{ 0.8, 0.05, 0.05, 0.05, 0.04, 0.01 \}$$

Le calcul de $U_1 (\sigma)$ pour $\sigma = 1, 2$ et 3 donne :

$$U_1 (1) = 0.014$$

$$U_1 (2) = 0.013$$

$$U_1 (3) = 0.014$$

$\Rightarrow U_1 (2) < U_1 (\sigma) \Rightarrow U_1 (\sigma)$ n'est pas une fonction décroissante de σ

III.4.3 - $U_2 (\sigma)$ vérifie la première propriété

$$\cdot H(Q) = H_M(Q) = \text{Log} (m - \sigma) \Rightarrow U_2(\sigma) = U_{2m}(\sigma) = 0$$

$$\cdot H(Q) = H_m(Q) = 0 \Rightarrow U_2(\sigma) = U_{2M}(\sigma) = A \cdot [(\text{Log} (m - \sigma)) / \text{Log} m]$$

$$0 \leq U_2(\sigma) \leq A \cdot [(\text{Log} (m - \sigma)) / \text{Log} m]$$

Comme $U_1 (\sigma)$, $U_2 (\sigma)$ est une fonction strictement décroissante de $H(Q)$ (cf annexe III.4.1)

III.4.4 - $U_2 (\sigma)$ vérifie la deuxième propriété :

A cet effet, nous allons montrer que $U_2 (\sigma) - U_2 (\sigma+1) = \Delta U_2 \geq 0$

$$\Delta U_2 = A \cdot [(\text{Log } (m-\sigma) - H(\sigma)) / \text{Log } m] - B \cdot [(\text{Log } (m-\sigma-1) - H(\sigma+1)) / \text{Log } m]$$

avec $\text{Log } m > 0, m > 1$

$$A \text{Log } (m-\sigma) - A H(\sigma) \stackrel{?}{\geq} B \text{Log } (m-\sigma-1) - B H(\sigma+1) \text{ ou bien :}$$

$$A \text{Log } (m-\sigma) - B \text{Log } (m-\sigma-1) \stackrel{?}{\geq} A H(\sigma) - B H(\sigma+1) \text{ or :}$$

$$A H(\sigma) - B H(\sigma+1) = P_{\sigma+1/i} \text{Log } [1 + (B / P_{\sigma+1/i})] + B \text{Log } [1 + (P_{\sigma+1/i} / B)] > 0$$

$$B = A - P_{\sigma+1/i}$$

$$A \text{Log } [[(m-\sigma) / (m-\sigma-1)] \cdot (1/A)] + (A - P_{\sigma+1/i}) \text{Log } (A - P_{\sigma+1/i}) \\ + P_{\sigma+1/i} \text{Log } P_{\sigma+1/i} (m-\sigma-1) \stackrel{?}{\geq} 0$$

or $\max P_{\sigma+1/i} = A$ et $\min P_{\sigma+1/i} = A / (m-\sigma)$, montrons que dans les deux cas, l'inégalité précédente est vérifiée.

. $P_{\sigma+1/i} = A$ (présence d'une seule raie) : dans ce cas :

$$P_{\sigma+1/i} \text{Log } ((m-\sigma) / (m-\sigma-1)) + P_{\sigma+1/i} \text{Log } (m-\sigma-1) = P_{\sigma+1/i} \text{Log } (m-\sigma) > 0$$

. $P_{\sigma+1/i} = A / (m-\sigma)$ (raies équiprobables), dans ce cas l'inégalité précédente est également vérifiée puisque par définition même de $U_2(\sigma)$ qui est nul lorsque les résidus sont répartis de façon équiprobable. Plus généralement, $\Delta U_2 \cdot \text{Log } m \geq 0$, en effet $A = \text{cte}, (m-\sigma) = \text{cte}$

$$\partial / \partial P_{\sigma+1/i} (\Delta U_2 \cdot \text{Log } m) = \text{Log } [(m-\sigma-1) (P_{\sigma+1/i}) / (A - P_{\sigma+1/i})] \Rightarrow$$

$$(\partial / \partial P_{\sigma+1/i}) (\Delta U_2 \cdot \text{Log } m) \geq 0 \text{ pour } P_{\sigma+1/i} \geq A / (m-\sigma) \text{ d'où :}$$

$$\Delta U_2 \text{ est } \geq 0 \quad \forall \quad A / (m-\sigma) \leq P_{\sigma+1/i} \leq A \quad \text{C.Q.F.D.}$$

III.4.5 - Quelques remarques importantes

. pour $\sigma = m-1, U_2(m-1) = 0$ or que $U_1(m-1)$ n'est pas définie

. si $H(Q) = H_m(Q) = 0 \Rightarrow U_2(\sigma) = A \cdot [\text{Log } (m-\sigma) / \text{Log } m]$ que devient $U_2(\sigma)$ lorsque σ/m est faible ou élevé ? Nous pouvons distinguer trois cas :

* m élevé et σ élevé (cas des règles inefficaces)

Dans cette situation $U_2(\sigma) \ll A$ et égale à 0 pour $\sigma = m-1$

Ceci semble logique, puisque au voisinage de m (avec m élevé) l'information résiduelle ne peut être opérationnelle.

* m élevé et σ faible (cas des règles efficaces)

$U_2(\sigma) \sim A$ puisque :

$$\text{Log}(m-\sigma) / \text{Log } m = [1 + \text{Log}(1 - (\sigma/m)) / \text{Log } m] \xrightarrow{\sigma \rightarrow m} 1$$

* σ faible et m faible (systèmes de faible dimension avec des variables à modalités réduites), σ n'étant pas négligeable devant m, alors :

$$U_2(\sigma) = A [\text{Log}(m-\sigma) / \text{Log } m] < A$$

C'est seulement dans ce cas que $U_2(\sigma)$ présente un inconvénient qui tend à limiter l'utilité de l'information perdue. Mais dans le cas général, les systèmes sont de grande dimension et les variables nécessaires à leur description sont importantes et présentent des modalités multiples d'où m est élevé. Par ailleurs, on souhaite extraire des règles précises et efficaces d'où la nécessité de choisir un σ aussi faible que possible.

ANNEXE III.5

L'ANALYSE FACTORIELLE DES CORRESPONDANCES

III.5.1. Notations et rappel des résultats de l'A.F.C.

. Dans l'espace \mathbb{R}^n , m points Y_j sont représentés, munis chacun d'un poids P_j et ayant pour coordonnées : $y_{ij} = P_{ij} / (P_j \sqrt{P_i})$

. Dans l'espace \mathbb{R}^m , n points X_i sont représentés, munis chacun d'un poids P_i et ayant pour coordonnées : $x_{ij} = P_{ij} / (P_i \sqrt{P_j})$

. (X_i) et (Y_j) désignent respectivement les nuages des points X_i et Y_j munis des

masses P_i et P_j . L'A.F.C. conduit aux résultats suivants :

- Le centre de gravité G de (Y_j) s'écrit : $G = \sum_j P_j Y_j$ avec $g_i = \sum_j P_j y_{ij} = \sqrt{P_i}$, où g_i est la $i^{\text{ème}}$ coordonnée de G .

- Le centre de gravité H de (X_i) s'écrit : $H = \sum_i P_i X_i$ avec $h_j = \sum_i P_i x_{ij} = \sqrt{P_j}$ où h_j est la $j^{\text{ème}}$ coordonnée de H .

- La variance est la même dans les deux cas puisque i et j jouent un rôle symétrique.

- La variance de la projection du nuage (Y_j) sur l'axe i est égale à la contribution de la ligne i à la valeur du χ^2/L et donnée par :

$$\sum_j P_j (y_{ij} - g_i)^2 = \sum_j [(P_{ij} - P_i P_j)^2 / (P_i P_j)]$$

- L'A.F.C. est basée sur la recherche des directions qui expliquent la plus grande partie de cette variance.

- Il existe des relations de passage et de dualité entre $[P_{ij}]$, (X_i) et (Y_j) , la connaissance d'un seul des trois éléments conduit à la détermination des deux autres.

- A un coefficient près, la projection d'un point du nuage dual sur un axe factoriel est la projection du barycentre des points du nuage direct pondérés par l'importance relative de la rubrique "duale" dans chacun d'eux.

- L'inertie expliquée par sous-espace à p dimensions est maximale lorsque celui-ci contient les vecteurs propres correspondant aux p plus grandes valeurs propres et égale à $\sum_1^p \lambda_a$.

La projection d'inertie expliquée par le sous espace p est $(\sum_1^p \lambda_a) / (\sum_1^n \lambda_a)$ où $\sum_1^n \lambda_a$ est l'inertie totale du nuage (Y_j) .

Quelques connaissances peut-on extraire de ces résultats ? Leur exploitation devient intéressante lorsqu'il est possible d'expliquer une grande partie de l'inertie du nuage à partir d'un sous-espace p de faible dimension. Ceci nous conduit à envisager l'analyse à partir d'un tableau de contingence $[P_{ij}]$ approximé.

III.5.2. Approximation de $[P_{ij}]$

On note Z_j la projection de (Y_j) sur un axe factoriel U correspondant à la valeur propre

$$\lambda, Z_j = (G + s_j U).$$

Z_j muni de la masse P_j définit un nuage (Z_j) auquel il est possible d'attribuer un tableau de fréquences $[\phi_{ij}]$. De façon symétrique, il est également possible de définir un nuage (R_i) muni de la masse P_i avec $R_i = H + r_i W$. La définition de Z_j permet la conservation des distributions marginales : $\phi_{i.} = P_i$ et $\phi_{.j} = P_j$ d'où :

$$\phi_{ij} / (P_j \sqrt{P_i}) = \sqrt{P_i} + s_j u_i \Rightarrow \phi_{ij} = P_i P_j (1 + (1/\sqrt{\lambda}) s_j u_i)$$

De façon générale, si l'on projette le nuage sur les p axes U_1, \dots, U_p , et W_1, \dots, W_p relatifs aux p plus grandes valeurs propres, on montre alors [VOL 76] que les fréquences relatives correspondant aux nuages des points projections sont :

$$\phi_{ij}^{(p)} = P_i P_j (1 + \sum_{k=1}^p (1/\sqrt{\lambda_k}) s_{jk} r_{ik})$$

avec s_{jk} : la coordonnée de Y_j sur U_k
 r_{ik} : la coordonnée de X_i sur W_k

Cette équation peut s'écrire également :

$$\phi_{j/i}^{(p)} = \phi_{ij}^{(p)} / P_i = P_j (1 + \sum 1/\sqrt{\lambda_k} s_{jk} r_{ik})$$

La suite des tableaux $[\phi_{ij}]^p$ constitue une suite d'approximations d'autant plus précises que p est élevé.

$$\text{Lorsque } p = n - 1 \text{ alors } [\phi_{ij}]^{n-1} = [P_{ij}]$$

ANNEXE IV

ANNEXE IV.1

Il s'agit de démontrer que $\Delta\chi^2 = 2 \text{ Log } 2 \cdot \Delta H$

$$\begin{aligned} \Delta H &= \sum_{j=1}^m [P_{ij} \text{ Log } [(P_{ij}/P_i) ((P_i+P_k) / (P_{ij}+P_{kj}))] \\ &\quad + P_{kj} \text{ Log } [(P_{kj}/P_k) ((P_i+P_k)/(P_{ij}+P_{kj}))]] \\ &= \underbrace{\sum_1^m P_{ij} \text{ Log } (P_{ij} / (P_i P_{.j}))}_{(1)} + \underbrace{\sum_1^m P_{kj} \text{ Log } (P_{kj} / (P_k P_{.j}))}_{(2)} \\ &\quad - \underbrace{\sum_1^m (P_{ij}+P_{kj}) \text{ Log } [(P_{ij}+P_{kj}) / (P_i+P_k)]}_{(3)} \end{aligned}$$

Montrons que si les distributions $P(Y/X = \alpha_i)$, $P(Y/X = \alpha_k)$ et $P(Y)$ sont voisines alors ΔH et $\Delta\chi^2$ sont identiques à un facteur près.

Posons $\text{Log}_2 X = \text{Log } X / \text{Log } 2$ et $X = (P_{ij}/P_{.j}) - 1$, en effectuant un développement limité de $\text{Log}(1+X)$ au voisinage de $X = 0$ et en s'arrêtant à l'ordre 2 on peut écrire que :

$\text{Log}(1+X) \cong X - X^2/2$, ainsi la quantité (1) s'écrit alors :

$$(1) = \sum_1^m P_i P_{.j} (1+X) ((X - (X^2/2))) = 1/(\text{Log } 2) \sum_1^m P_i P_{.j} (X + (X^2/2) - (X^3/2)),$$

en négligeant le terme de 3^e ordre et sachant que $\sum_1^m P_i P_{.j} X = \sum_1^m (P_{ij} - P_i P_{.j}) = P_i - P_i \cong 0$, (1) devient :

$$1/(\text{Log } 2) \sum_1^m P_i P_{.j} ((X^2/2)) = 1/(2 \text{ Log } 2) \sum_1^m P_i P_{.j} ((P_{ij}/P_{.j}) - 1)^2$$

En procédant de la même façon que (1), (2) et (3) s'écrivent respectivement :

$$(2) = 1/(2 \text{ Log } 2) \sum_1^m P_k P_{.j} ((P_{kj}/P_{.j}) - 1)^2$$

$$\begin{aligned}
(3) &= -1/(2 \text{Log } 2) \sum_1^m (P_i + P_k) P_j [((P_{ij} + P_{kj}) / ((P_i + P_k) P_j)) - 1]^2 \\
2 \text{Log}_2 \Delta H &= \sum_1^m [P_k P_j ((P_{jk} / P_j) - 1)^2 + P_i P_j ((P_{ji} / P_j) - 1)^2 \\
&\quad - (P_i + P_k) P_j (((P_{ij} + P_{kj}) / ((P_i + P_k) P_j)) - 1)^2] \\
&= \sum_1^m [((P_{ij} - (P_i P_j))^2 / P_i P_j) + ((P_{kj} - (P_k P_j))^2 / P_k P_j) \\
&\quad - ((P_{ij} + P_{kj} - (P_i + P_k) P_j)^2 / ((P_i + P_k) P_j))] \\
&= \sum_1^m [P_k (P_i + P_k) (P_{ij} - P_i P_j)^2 + P_i (P_i + P_k) (P_{kj} - P_k P_j)^2 - P_i P_k (P_{ij} + P_{kj} \\
&\quad - (P_i + P_k) P_j)^2] / [(P_i + P_k) P_i P_k P_j] \\
&= \sum_1^m \{ [P_i^2 P_k^2 / (P_i + P_k) P_j] [(P_{ji} - P_{jk})^2 / P_i P_k] \} \\
&= [P_i P_k / (P_i + P_k)] \sum_1^m [(P_{ji} - P_{jk})^2 / P_j] = \Delta \chi^2 \\
2 \text{Log } 2 \Delta H &= \Delta \chi^2 \quad \text{C.Q.F.D.}
\end{aligned}$$

ANNEXE IV.2

LES PROPRIETES FONDAMENTALES DE L'ALGEBRE DE POST

Définition 1 : Soit $m \geq 2$. Une algèbre de POST est un treillis distributif P avec un élément nul 0 et un élément universel U pour lequel les axiomes suivants sont vérifiés :

axiome 1 : P possède n éléments e_0, e_1, \dots, e_{n-1} tels que :

- a) $0 = e_0 < e_1 < \dots, e_{n-1} = U$
- b) si $X, e_i \in P$ et $X \cdot e_i = 0$ avec $i \neq 0$ alors $X = 0$
- c) si $X, e_i, e_j \in P$ et $X + e_i = \bar{e}_j$ pour $i < j$ alors $X = e_j$

axiome 2 : $\forall X \in P$, il existe une suite d'opérateurs unaires sur X notés

$[C_0(X), C_1(X), \dots, C_{n-1}(X)] \in P$ tels que pour tout

$i, j \in [0, 1, \dots, n-1]$

$$a) C_i(X) \cdot C_j(X) = 0 \text{ pour } i \neq j$$

$$b) \sum_{i=0}^{n-1} C_i(X) = U$$

$$c) X = \sum_{i=1}^{n-1} C_i(X) \cdot e_i$$

Définition 2 : Soient X et $X' \in P$

X' est le complément de X dans P ssi :

$$X \cdot X' = 0$$

$$X + X' = U$$

Tous les éléments d'une algèbre de POST n'admettent pas de complément, mais dans toute algèbre de POST, il existe au moins deux éléments qui admettent un complément : 0 et U.

Définition 3 : soit $X \in P$, l'opérateur $d_i(X)$ est défini par :

$$d_i(X) = \begin{cases} 0 & \text{si } X \neq i \\ U & \text{si } X = i \end{cases} \quad i \in \{0, 1, \dots, n-1\}$$

Cet opérateur est représenté par :

$X(0 \dots 0 U 0 \dots 0)$ où $(0 \dots 0 U 0 \dots 0)$ est un n-uple de 0 sauf pour la $i^{\text{ème}}$ composante égale à U.

Les fonctions de POST

Les fonctions de POST jouent le même rôle dans les algèbres de POST que les fonctions booléennes dans l'algèbre de BOLLE et peuvent s'écrire sous forme canonique.

Définition 4 :

Une fonction de POST de p variables X_0, X_1, \dots, X_{p-1} ou $X_j \in P$ et $j = 0, 1, \dots, p-1$ est une fonction qui peut être construite à partir :

. des fonctions constantes :

$$\varepsilon_i(X_0, X_1, \dots, X_{p-1}) = e_i, \quad i = 0, 1, \dots, p-1$$

. des fonctions identités :

$$I_j (X_0, X_1, \dots, X_{p-1}) = X_j, j = 0, 1, \dots, p-1$$

. d'un nombre fini d'opérateurs :

$$+, \cdot, C_0, C_1, \dots, C_{n-1}$$

Théorème :

Si f est une fonction de POST de p variables X_0, X_1, \dots, X_{p-1} alors :

$$f(X_0, X_1, \dots, X_{p-1}) = \sum_{\substack{i_j=0 \\ j=0}}^{n-1} f(e_{i_0}, e_{i_1}, \dots, e_{i_{p-1}}) \cdot C_{i_0}(X_0) \cdot \dots \cdot C_{i_{p-1}}(X_{p-1})$$

Pour plus de détails cf à [WOJCIECHOWSKI 79] [ETIEMBLE 79].

BIBLIOGRAPHIE

Z. ABID

"Contribution à l'analyse structurale par la théorie des demi-treillis"
Thèse de 3^{ème} Cycle, Université Claude Bernard, Lyon 1 ; 1979.

N.L. AGGRWAL

"Mesures d'information : caractéristiques et propriétés"
Ecole de l'INRIA :theorie de l'information 1976

W.R. ASHBY

"Introduction to cybernetics"
CHAMPAN and Hall ,London,1956

W.R. ASHBY

"Measuring the internal informational exchange in a system"
Cybernetica Namr,Belgique,vol 8,1965

W.R. ASHBY

"Two tables of identities governing information flows within large systems"
Communication of the American society of cybernetics ,vol 1,1969

P.AYGALINC

"Application de la reconnaissance des formes à l'aide au diagnostic médical, sélection multi-critère des variables explicatives"
Thèse de doctorat ,université de Lille 1,1986

P. AYGALINC, M. STAROSWIECKI, O. DUBOIS

"Selection of explicative variables for aided diagnosis"
Second catalan international symposium on statistics ,Barcelone,sep ,1986

M. BARBOUCHA

"Contribution à l'automatisation d'un bassin biologique de traitement des eaux usees"
Thèse de troisième cycle, université de Lille 1, juin 1981

M. BARBOUCHA , R. LAURENT

"Sur la conduite des processus complexes:application aux stations d'épuration des eaux"
Article paru dans l'ouvrage de J.C PRUVOST "point en Automatique",éditions technique et documentation, PARIS 1982

M. BARBOUCHA ,R. LAURENT

"Comparaison of three regulation methods for particular incompletely controlable multivariable systems"
Int Jour of systems science,vol 10,N°1,1984

- M. BARBOUCHA, M. STAROSWIECKI**
 "Modelisation structurale des systèmes complexes"
Sixth inter symposium modelling, identification and control, IASTED, Grindelwald Suisse, february 1987
- M. BARBOUCHA, M. STAROSWIECKI**
 "Extraction des connaissances à partir d'un tableau de contingence"
Fifth inter symposium applied informatics, IASTED, Grindelwald Suisse february 1987
- M. BARBOUCHA, M. STAROSWIECKI, P. AYGALINC**
 "Génération des règles à partir d'un tableau de contingence"
Article accepté aux 5^o journées internationales: analyse des données et informatique, INRIA, 29 sep-2 oct 1987- Verssailles
- M. BARBOUCHA, M. STAROSWIECKI**
 "Knowledge extraction from contingency tables"
Article accepté à seond inter conf on human-computer interaction, august, Honolulu HAWAII
- J.P. BENZECRI**
 "Annalyse des données"
DUNOD 1980
- P. BERTIER, P.M. BOUROCHE**
 "Analysé des données multi-dimensionnelles"
Presse universitaire de France 1975
- J.C. BERTRAND, J.J. MERCIER, N. GIAMBIASI**
 "Sur la recherche de l'inverse d'un circuit combinatoire"
RAIRO, juillet 1974
- G. BOSSUET**
 "Etude d'un modèle algébrique permettant une matérialisation optimale de l'opérateur table de vérite sur un ensemble de fonctios associatives multi-valuées"
Thèse d'état, Université Paris 6, janvier 1977
- M. BOURTON**
 "Contribution à l'analyse, le traitement et la reconnaissance des formes ponctuelles: application à la classification des couches géologiques"
Thèse d'état, université de Lille 1, 1975
- R. BRADDOCK, G. EPSTEIN, H. YAMANAKA**
 "Multiple valued logic design and applications in binary computers"
71 ISMVL pp13-25
- F. CAILLEZ, J. P. PAGES**
 "Introduction à l'analyse des données "
Société de mathématiques appliquées et sciences humaines, 1976
- R.C. CONANT**
 "The information transfer required in regulatory processes"
IEEE Trans. on Systems, Sciences and Cybernetics, vol. SSC-5, n°4, 1969

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, 1970

R.C. CONANT

"Detecting subsystems of a complex system"
IEEE Trans. on Systems Man and Cybernetics, vol. SMC-2, n°4, 1972

R.C. CONANT

"Laws of information which govern systems"
IEEE Trans on Systems Man and Cybernetics, vol. SMC-6, n°4, 1976

M.O. CORDIER, M.C. ROUSSET

"Le contrôle dans le moteur d'inférence"
*5^{ème} journées internationales des systemes experts et leurs applications, Avignon
Mai 1985*

M.COURVOISIER, R. VALETTE

"Commande des procédés continus : logique séquentielle"
Dunod Université, 1986

DACHUNA-CASTELLE, DUFLO

"Probabilités et statistiques"
Tome 1, Problème à temps fixe, Masson 1982

E. DEMONCHAUX, J. QUINQUETON, H. RALAMBONDRAIN

"CLAVECIN un système expert en analyse des données"
Rapport de recherches INRIA, n° 431, 1985

D. DUBOIS, H. PRADE

"Théorie des possibilités : applications à la représentation des connaissances en informatique"
Editions MASSON, Paris 1985

B. DUBUISSON, P. LAVISON

"Surveillance of a nuclear reactor by use of a pattern recognition methodology"
IEEE Trans. on Systems Man and Cybernetics, vol. SMC-10, n°10, october, 1980

R.O DUDA & Col

"Model designed in the PROSPECTOR consultant system for mineral exploitation"
Expert systems in the micro-electronic age, Editions MICHIE, Edinbourg, 1980

J. DUFOUR, G. GILLES, C. FOULARD

"Analyse structurale et partition des systemes dynamiques complexes à l' aide de la théorie de l' information. Etude de l'influence du nombre de classes"
Comptes rendus de l'académie des sciences de Paris, t. 282, série A, Mars 1976

J. DUFOUR, G. GILLES

"A methodology for structural analysis and partition of dynamic systems. Application to two macroeconomic systems"
2^{ème} congrès IFAC Helsinki, 1978, publié par pergamon press

J. DUFOUR

"Méthodes et méthodologies d'analyse de systèmes complexes. Application aux procédés industriels et aux systèmes macroéconomiques"
Thèse d'état, Univ. Claude Bernard, Lyon, Mai 1979

J. DUSSAULT, G. METZE, M. KRIEGER

"A multivalued switching algebra with boolean proprieties"
ISMVL, pp 68-73, 1976

A. DUSSAUCHOY

"Generalized information theory on some ordered sets and structure of systems"
14th Annual conf. on Information Sciences in Systems, Princeton University, 1980

S. ELMALEH, P. BEN AIM

"Représentation de l'écoulement macroscopique dans un bassin d'aération par une cascade de deux mélangeurs parfaits inégaux"
The chemical engineering journal, 9, 1975

D. ETINEMBLE

"Contribution à l'étude des circuits multivalués et leurs utilisations"
Thèse d'état, Université Paris 6, Juin 1979

W.R. GARNER, N.J. MCGILL

"The relation between information and variance analysis"
Psychometrika, vol. 21, 1956

S.GENTIL

"Analyse en composantes principales du modèle d'un écosystème"
RAIRO automatique, vol. 18, n°4, 1984

M.GHOUALMI

"Analyse et identification de l'écoulement dans un bassin d'épuration des eaux, connexions avec les modèles biologiques"
Thèse de Docteur Ingénieur Université de Lille 1, octobre 1978

C. GRANGER

"Reconnaissance d'objets par mise en correspondance en vision par ordinateur"
Thèse de doctorat en informatique, Université de Nice, Juin 1985

S. GUIASU

"information theory with applications"
MC-GRAW Hill, London, 1977

M.HAMAD

"Validation des mesures et détection des capteurs défaillants dans un système de contrôle-commande"
Thèse de doctorat, Université de Lille 1, 1986

R.L. HERMANN

"Selection and implementation of a ternary switching algebra"
AFIPS Conference. Prod, vol 32, pp 283-290, 1970

- M. JAMBU**
 "Classification automatique pour l'analyse des données"
Dunod, 1978
- G. JUMARIE**
 "Théorie relativiste de l'information et télécommunications, perspectives"
Ann. Télécom., 33, n°1-2, Janvier 1978
- G.J. KLIR**
 "An approach to general systems theory"
Van Nostrand Reinhold, New-York, 1969
- G. KLIR**
 "Identification of generative structures in empirical data"
International journal of general systems, vol. 3, 1976
- G. KLIR**
 "On the representation of activity arrays"
International journal of general systems, vol 2, 1975
- G. KLIR**
 "On the problem of computer-aided structure identification : some experimental observations and resulting guidelines"
International journal of Man-Machine Studies, vol 9, 1977
- Y. KODRATOFF, J.G. GANASCIA**
 "Démonstration automatique de théorèmes et systèmes experts, quelques faits élémentaires"
5^{ème} journées internationales des systèmes experts et leurs applications, Avignon-Mai 1985
- S. KULLBACK**
 "Information theory and statistics"
Dover publications, NEW YORK 1962
- A.H. LABORDERE**
 "Analyse des données : applications et méthodes pratiques"
Editions Masson 1977
- P. LARMINAT, Y. THOMAS**
 "Automatique des systèmes linéaires"
Tome 1 : Signaux et systèmes, Flammarion Sciences, 1975
- J.P. LAURENT**
 "La structure de contrôle dans les systèmes experts"
TSI, vol 3, n° 3, 1984
- R. LAURENT, M. BARBOUCHA**
 "Control system for a wastewater biological purification process"
Conf. Inter. on Systems Engineering , COVENTRY, ENGLAND, sept. 1980
- R. LAURENT, M. BARBOUCHA, L. POVY**
 "Conception d'un système de surveillance d'une station d'épuration des eaux usées"
International journal of modelling and simulation, vol 1, n°4, 1981

- R. LAURENT**
 "Modélisation et identification des systèmes complexes, application à l'épuration biologique des eaux"
Thèse d'état, Université de Lille 1, Mars 1985
- J.L. LAURIERE**
 "Représentation et utilisation des connaissances"
TSI, vol 1, n° 1, 1982
- P. LEBEUX, D. FONTAINE**
 "Un système d'acquisition des connaissances pour systèmes experts"
TSI, vol 5, n° 1, 1986
- Y. LECHEVALLIER**
 "Classification automatique optimale sous contrainte d'ordre total"
INRIA, rapport de recherche, n° 200, 1976
- A. LELAN**
 "Etude des cuves mécaniquement agitées"
Thèse d'état, Grenoble, 1972
- I.C. LERMAN**
 "Reconnaissance et classification des structures finies en analyse des données"
Théorie et Méthodes, vol.1, 1976
- I.C. LERMAN**
 "Classification et analyse ordinale des données"
Dunod, Paris 1981
- R. LOPEZ DE MANTARAS BADIA**
 "Auto-apprentissage d'une partition. Application au classement itératif de données multidimensionnelles"
Thèse de doctorat de spécialité, Université Paul Sabatier, Toulouse, 1977
- A. MALLOUM**
 "Analyse structurale et partition des systèmes macroéconomiques dynamiques complexes, application à MINI-DMS"
Thèse de 3ème cycle, Lyon 1980
- C. MARFAING PALLIER**
 "Contribution à l'analyse structurale des systèmes complexes par le biais des concepts de la théorie de l'information"
Thèse de docteur ingénieur, INSA, Toulouse, décembre, 1983
- M. MARRAKCHI, B. HOURIEZ, F. GRZESIAK, D. WILLAYES**
 "Un système expert d'aide au diagnostic"
5ème journées Internationales des systèmes experts et leurs applications, Avignon, Mai, 1985
- R. MARTIN-CLOUAIRE**
 "ELFIN : Une approche système expert et théorie des possibilités appliquée en géologie pétrolière"
Thèse 3ème cycle, Université Paul Sabatier, Toulouse 1984

J.J. MERCIER

"Optimisation des structures auto-testables D-invertibles"
RAIRO automatique, vol 11, n°2, 1977

M. MILGRAM

"Méthode de décomposition de systèmes"
Thèse de 3^{ème} cycle, Université de Technologie de Compiègne, 1975

G.A. MILLER

"What is information measurement"
American Psychologist, vol. 8, n°2, 1963

R.S. NUTTER, R.E. SWARTWOUT, D.C. RINE

"Equivalence and transformation for POST multivalued algebras"
IEEE TC, vol. C23, March, 1974

G. OSWALD

"Théorie de l'information ou analyse diacritique des systèmes"
Editions Masson, Paris, 1986

E. PARZEN

"Modern probability theory and its application"
John Wiley and sons, New York, 1960

G. PASS

"Consistent evaluation of uncertain reasoning systems"
6^{ème} journées internationales des systèmes experts et leurs applications, Avignon, Avril, 1986

M. PELEGRIN

"Automatique et grands systèmes"
Congrès automatique, AFCET, Nantes, 1981

C.F. PICARD

"Graphes et questionnaires"
Editions Gauthier Villars, Paris, 1972

C.F. PICARD

"Chronique sur la théorie de l'information"
RAIRO informatique théorique, vol. 2, n°12, 1978

S. PINSON

"Représentation des connaissances dans les systèmes experts"
RAIRO informatique, vol. 15, n°4, 1981

H. PRADE

"Modèles mathématiques de l'imprécis et de l'incertain en vue d'application au raisonnement naturel"
Thèse d'état, Université Paul Sabatier, Toulouse, 1982

- B. PRADIN**
 "Méthodes et outils logiciels d'analyse structurale au moyen de la théorie de l'information"
APII, 20, 317-330, 1986
- PRESTO, A. HUGER**
 "La programmation des réseaux logiques"
ETI, n° 22, Juin 1985
- V. RIALLE**
 "Génération de connaissances expertes : utilisation de l'analyse des données"
5^{ème} Journées Internationales des systèmes experts et leurs applications, Avignon, Mai, 1985
- M. RICHETIN**
 "Analyse structurale des systèmes complexes en vue d'une commande hiérarchisée"
Thèse d'état, Université Paul Sabatier, Toulouse, 1975
- R.E. RINK**
 "Infortion theoretic methods for modelling and analysing large systems"
5th IFIP conference on Opt. Tech., Rome, 1973
- E. ROUBINE**
 "Introduction à la théorie de la communication "
Tome 3, théorie de l'information, Masson 1970
- J. SALLATIN**
 "Représentations d'observations dans le contexte de la théorie de l'information"
Thèse d'état, Université Pierre et Marie Curie, Juin 1979
- G. SAPORTA**
 "Théorie et méthode de la statistique"
Paris-Technip. 1978
- M. SBAI**
 "Analyse structurale des systèmes complexes : méthodes d'explication et de partition"
Thèse de 3^{ème} cycle, Université de Lille 1, 1983
- G. SHAFER**
 "A mathematical theory of evidence"
Priceton University Press, 1976
- C.E. SHANNON**
 "A mathematical theory of communication"
Bell system technical journal, vol. 27, 1948
- E.H. SHORTLIFF**
 "Computer-based medical consultation, : MYCIN"
American Elsevier, 1976
- H.A. SIMON**
 "The architecture of complexity"
Proc. of the american philosophical society, vol. 106, n° 6, 1962

- D. SLEPIAN**
 "Key papers in the development of information theory"
IEEE Press, New York, ref. PC00299, 1974
- G. SOULA**
 "Aide à la décision en logique floue, application en médecine"
Thèse d'état, Faculté de médecine, Université d'Aix-Marseille, 1981
- M. STAROSWIECKI**
 "Contribution à l'analyse et à la commande de systèmes complexes à critères multiples"
Thèse d'état, Université de Lille 1, Janvier, 1979
- M. STAROSWIECKI, V. TORO**
 "Structural analysis of complex systems by means of canonical analysis"
VII International conf. on systems science, Wroclaw, Pologne, sept. 1981
- M. STAROSWIECKI, M. BARBOUCHA**
 "Modélisation structurale des systèmes complexes"
6^{ème} congrès internationale de la WOGSC, Paris, sept. 1984
- M. STAROSWIECKI**
 "Analyse structurale des systèmes complexes"
RAIRO automatique, vol.18, n°2, 1984
- M. STAROSWIECKI, P. AYGALINC, O. DUBOIS**
 "Sélection de variables explicatives et prescription d'examens complémentaires pour une aide à la décision médicale"
Conf. Internationale. IPMU, Paris, Juin-Juillet, 1986
- M. STAROSWIECKI, M. HAMAD**
 "Validation of measurement and detection of sensors' failures in control systems"
EUSIPCO-86, The Hague, The Netherlands, sept. 1986
- M. SUGENO**
 "Theory of fuzzy integral and its applications"
Ph.D Thesis, Tokyo Inst. of Technology, Japon, 197
- P. TALLIBERT**
 "M13 outils de représentation et d'exploitation de la connaissance et exemple d'utilisation"
5^{ème} journées internationales des systèmes experts et leurs applications, Avignon, Mai, 1985
- M. TERRENOIRE, D. TOUNISSOUS**
 "Pseudo-questionnaires et informations"
Decision making and medicinal care, North Holland Publishing Company, 1976
- V. TORO, M. STAROSWIECKI**
 "Méthodes heuristiques et décomposition optimale des systèmes complexes"
Congrès AFCET automatique, Nantes, oct. 1981

- V. TORO DE CORDOBA**
 "Contribution à l'analyse structurale des systèmes complexes à l'aide de l'entropie et ses généralisations"
Thèse 3ème cycle, Univ. de Lille 1, Mars, 1982
- T. TRACZYCK**
 "Axioms and some properties of POST algebras"
Coll. math, vol. 10, pp 193-209, 1963
- M. VOLLE**
 "Analyse des données"
Publications de l'institut nationale de la statistique et des études économiques, 1976
- M. VOLLE**
 "Analyse des données"
2ème édition economics, 1982
- T.J. WAGNER**
 "Non parametric estimates of probability densities"
IEEE Trans. on information theory, vol. IT-21, n° 4, 1975
- B. WALLISER**
 "Systèmes et modèles. Introduction critique à l'analyse de systèmes"
Edition du seuil, Paris, 1977
- W.S. WOJCIECHOWSKI**
 "Multiple-valued combinational logic design using theorem proving"
Ph.D Thesis, Illinois Institute of Technology, Chicago, Illinois, 1979
- A.D WYNER**
 "Fundamental limits in information theory"
Proc. of IEEE, vol. 69, n° 2, Feb. 1981
- YASUICHI HORIBE**
 "Entropy and correlation"
IEEE Trans. on System, Man and Cybernetics, vol. SMC 15, n° 5, 1985
- T.Y. YOUNG, P.S. LIU**
 "Linear transformation of binary random vector and its application to approximating probability distribution"
IEEE Trans. on Information Theory, vol. IT 24, n° 2, March 1978
- B. YVINEC, M. MILGRAM, B. DUBUISSON**
 "Estimation de l'entropie dans le cas continu en vue de la décomposition des systèmes complexes"
RAIRO automatique, vol 12, n° 1, 1978
- L.A. ZADEH**
 "Fuzzy sets as a basis for a theory of possibility"
Fuzzy sets and systems, n° 1, 3-28, 1978
- A. ZIGHED, M. TERRENOIRE, D. TOUNISSOUS**
 "Système interactif pour un processus d'interrogation non arborescent"
Journées de statistique, Montpellier, 21-24 Mai, 1984